RICE UNIVERSITY

Functional and Structural Studies of Influenza B Virus Hemagglutinin

by

Fengyun Ni

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Jianpeng Ma, Chair
Lodwick T. Bolin Professor of
Biochemistry
Baylor College of Medicine,
Professor of Bioengineering
Rice University

Yizhi Jane Tao
AssociateProfessor
Biochemistry and Cell Biology
Rice University

Kathleen Matthews
Stewart Memorial Professor
Biochemistry and Cell Biology
Rice University

HOUSTON, TEXAS
MAY 2013

RICE UNIVERSITY

Functional and Structural Studies of Influenza B Virus Hemagglutinin

by

Fengyun Ni

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Jianpeng Ma, Chair
Lodwick T. Bolin Professor of
Biochemistry
Baylor College of Medicine,
Professor of Bioengineering
Rice University

Yizhi Jane Tao
AssociateProfessor
Biochemistry and Cell Biology
Rice University

Kathleen Matthews
Stewart Memorial Professor
Biochemistry and Cell Biology
Rice University

HOUSTON, TEXAS
MAY 2013

ABSTRACT

Functional and Structural Studies of Influenza B Virus Hemagglutinin

by

Fengyun Ni

Influenza A and B viruses are major causes of seasonal flu epidemics each year. Hemagglutinin (HA) mediates the binding of virus to host cell and the fusion with host membrane. The crystal of HA in complex with antibody that reveals the mechanism by which antibody recognizes HA may not diffract to high resolution, thereby preventing the accurate interpretation of the structural model. The application of normal mode refinement that aims for improving the structure quality at the low resolution is tested. These studies provide some guidelines for future refinement of HA-antibody complex structures. By comparing the residues constituting the base of the receptor binding site of influenza A and B virus HAs, it is found that they share some similarities, except for a Phe at position 95 of influenza B virus hemagglutinin (BHA) versus Tyr in of influenza A virus hemagglutinin (AHA). The recombinant protein BHA containing the F95Y mutation exhibits the increased receptor binding affinity and specificity. However, recombinant viruses with the Phe95→Tyr mutation show lower erythrocyte agglutination titer and decreased binding abilities with different cell lines. The replication of the Phe95→Tyr mutant virus in mice is also attenuated. These data suggest that the increased receptor binding ability of HA alone is not advantageous to the pathogenesis of the viruses.

The structure of BHA_2 (a portion of BHA near the C-terminus) at the post-fusion state has been determined to 2.45 Å resolution. This protein forms a hairpin-like conformation rich in α-helices. About 70 residues from the N-terminus is a three-stranded coiled coil, and the remaining of the protein packs in anti-parallel against the groove formed by the central helices. In the post-fusion state of BHA_2, the helix converted from the B-loop in pre-fusion state contacts the C-terminal fragment of this protein with more hydrophobic interactions as compared to AHA_2. This structure illustrates the distinct stabilization strategy employed by BHA_2 to form a post-fusion state that resembles that for AHA_2. These studies will further the understanding of BHA with respect to its role in receptor binding ability and fusion.

Acknowledgments

I would like to thank my great advisor, Dr. Jianpeng Ma, for his guidance and assistance. His passion for science touches deeply in my heart and lets me know what I should pursue as a graduate student. He always keeps thinking and dreaming of achieving much higher goals for research, which encourages me so much that I feel confident of the future when I meet difficulties. He also teaches me a lot that I should take the responsibility of each role that I am playing in the life.

I also want to thank Dr. Qinghua Wang. She let me know the importance of communicating my ideas and experiences with others. Her wide knowledge of both biological science and crystallography impresses me, and I have learned many things from her. She is energetic in several different aspects of research and opens the wonderful scientific world for us all in the lab.

I'd like to thank my committee members, Dr. Kathleen Matthews and Dr. Yizhi Jane Tao. The communications with them let me learn many things about a good scientist.

I also want to thank Dr. Bruce Johnson, Dr. Yvonne Kambourelis and Dr. Carolyn Aresu. They helped me with study and life in Rice. Carolyn helps a lot on the paperwork to proceed as a PhD student.

I also want to thank my lab fellows, Dr. Billy Poon, Dr. Mingyang Lu, Dr. Cheng Zhang, Dr. Xiaorui Chen, Jialin Li, Dr. Athanasios Dousis, Chaoping Qin, Zhao Ge, Dr. Brian Kirk, Yufeng Gou, Dr. Junqing Du, Dr. Jun Shen, and Patty Wu. They are my good friends, and all give me much help. With them, I could always discuss when I met problems in experiments.

I also want thank Dr. Lin Liu, Dr. Innocent Nnadi Mbawuike and Dr. Diane Niño at Baylor College of Medicine. They have given me advice and shared with me their experiences regarding the studies on influenza virus.

I want to say thank you to my dear parents and my dear grand-mother. Without them, I could do nothing. They are the foundations of my life. Though I have left them since I came Rice, my heart will always be with them.

Above all, I want to thank my beloved wife, Fei Zheng, for her support and understanding. She always stands by my side whenever I need her assistance. We are working together to get a brighter future. She brings me two little lovely daughters and one little cute son and makes our life colorful and full of happiness. I will cherish her and our children forever.

Contents

List of Figures

List of Tables

List of Equations

Abbreviations

HA	Hemagglutinin
BHA	Influenza B virus Hemagglutinin
AHA	Influenza A virus Hemagglutinin
SA	Sialic Acid
RBC	Red Blood Cell
NMRef	Normal Mode Refinement
TLS	Translation, libration and screw-rotation

Chapter 1

Introduction

1.1. Influenza virus

Three genera of influenza viruses (A, B, and C) are members of *Orthomyxoviridae* family (*1*). The viruses in the family of *Orthomyxoviridae* have a negative-sense, single-stranded, and segmented RNA genome (*1*). Human Influenza A and B viruses can cause disease in people of any age, and they remain as a leading cause of mortality and morbidity worldwide. Infections of Influenza C viruses cause a milder respiratory illness and are thought to be endemic (*2*). Health organizations including the U.S. Food and Drug Administration (FDA), World Health Organization (WHO), the U.S. Centers for Disease Control and Prevention (CDC) and other institutions are devoted to the identification of the most common influenza viruses during the upcoming flu season. The vaccine against those identified viruses are produced to protect people from potential infections of influenza viruses.

Influenza A, B, and C viruses evolved from a common ancestor, and it is proposed that influenza B viruses diverged from influenza C viruses about 8,000 years ago; influenza A viruses diverged from influenza B viruses about 4,000 years ago; and the different subtypes of influenza A viruses diverged about 2,000 years ago based on the sequence comparison of hemagglutinin (*3*). Evolutionary rates for influenza A viruses are much faster than influenza B and C viruses (*4*). It is postulated that the transmission of influenza A viruses to human happened less than 150 years ago, possibly from birds (*5*). Influenza A viruses have a wide host range, including humans, avian species, and domestic animals (*6*); influenza B viruses only infect humans, seals, and ferrets (*7*–*9*); and influenza C viruses only infect humans, pigs and dogs (*9*). Recent epidemic among New England harbor seals caused by a mutant avian H3N8 influenza A virus leads to concerns about transmission to human (*10*). The evolution of influenza A viruses has not reached an adaptive equilibrium, whereas influenza B and C viruses with limited host range are considered to have achieved an evolutionary stasis in human (*5*).

Despite the effort to develop vaccine against influenza virus, this virus is still able to elude host adaptive immunity by mutation of its genome, resulting in a new strain of influenza virus that cannot be recognized by the antibodies against the older strain of virus (*11*). The mutations mostly happen on the surface proteins of influenza viruses, hemagglutinin and neuraminidase, because these two proteins are the main antigens recognized by the antibodies. The continuous genetic change and antigenic change of influenza viruses is often referred as antigenic drift, and this process is common in both influenza A and B viruses (*12*, *13*). Another type of

change that only occurs in influenza A virus is called antigenic shift (*14*). The antigenic shift is a more dramatic and abrupt change due to the recombination of viral genetic materials from more than one original strain that results in a new subtype of influenza virus with a new combination of hemagglutinin and neuraminidase proteins (*15*).

Based on the antigenic properties of hemagglutinin (H) and neuraminidase (N) antigens, influenza A viruses are classified into different subtypes depending on the types of hemagglutinin and neuraminidase present on the virus surface (*1*, *4*). There are 16 distinct subtypes of hemagglutinin and 9 subtypes of neuraminidase in wild waterfowl. However, influenza B and C viruses do not have subtypes (*4*).

The nomenclature of influenza viruses includes the type of virus (A, B or C), the host of origin (e.g., swine, chicken, etc., omitted if human.), the place of isolation (e.g., Texas, California, etc.), the strain number, the year of isolation and subtypes (only for influenza A viruses) (*16*). For example, A/Victoria/361/2011 (H3N2), which is one of the vaccine viruses in 2012-2013 influenza season, stands for the 361st strain of an H3N2 subtype influenza A virus isolated from a human in Victoria in 2011.

1.2. Morphology and genome of influenza virus

As an enveloped virus, influenza virus derives its lipid bilayer from the plasma membrane of the host cell (Figure 1.1, grey circle). There are two glycoproteins on the surface of the virion, hemagglutinin (HA) and neuraminidase

(NA). The envelope also harbors the ion channel protein M2. A layer of the matrix protein 1 (M1) that forms an ordered structure lies adjacent to the lipid envelope. Each of the eight pieces of RNA segments is associated with the polymerase subunits (Polymerase Basic 1[PB1], Polymerase Basic 2 [PB2] and Polymerase Acid [PA]) and nucleoprotein (NP) oligomers in the form of ribonucleoprotein (RNP) that locates in the core of the virion (*17*). Nonstructural protein (NS2) also exists in the viral particle (*18*). Influenza viruses display multiple morphologies, including spherical viral particles with about 100 nm in diameter and elongated filamentous viral particles with about 100 nm in diameter and up to 20 μm in length (*19–21*).

Both influenza A and B viruses have eight RNA segments including PB2, PB1, PA, HA, NP, NA, M, and NS according to the segment sizes. The genome of influenza A virus encodes 12 proteins in total, whereas the genome of influenza B virus encodes 11 proteins in total. Near the 5' end of the second RNA PB1 of influenza A virus, an alternative open reading frame gives rise to the 87-amino acid-long PB1-F2 protein (*22*), which are not observed in influenza B virus. The PB1 RNA also encodes the N-terminal truncated form of protein N40 (*23*). The sixth RNA NA of influenza B virus also encodes the 100-amino acid-long NB protein (*24*), which does not exist in influenza A virus. The seventh RNA M of influenza A virus not only encodes the M1 protein, but also results in a spliced mRNA encoding the M2 protein (*1*). The corresponding RNA M of influenza B virus also encodes the M1 protein, but utilizes a distinct "stop-start" translation mechanism to translate the BM2 protein (*25*).

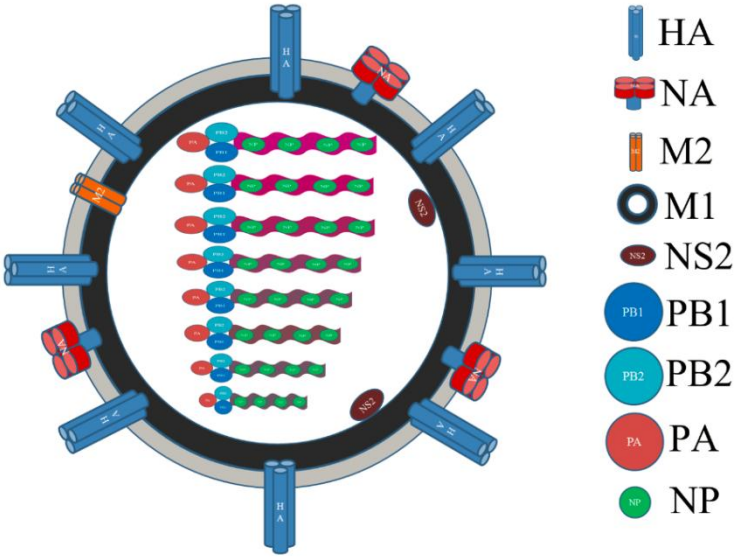

Figure 1.1 The generic structure of influenza virus. The lipid bilayer envelope of influenza virus is shown in grey, the layer of the M1 protein that lies under the envelope is shown in black. The protein components are shown in different shapes and colors as labelled in the figure. The viral ribonucleoproteins are shown in eight segmented pieces with different lengths.

1.3. The life cycle of influenza virus

The sialic acids (SAs) on the cell surface are the common receptors for influenza A and B viruses, where surface protein HAs of influenza viruses interact with the receptors (Figure 1.2). The terminal sialic acid is connected to the penultimate galactose (Gal) with either α(2,3)- or α(2,6)-linkage. Human tracheal epithelial cells are reported to have mainly SAα(2,6)Gal sugar moieties (*26*), while duck gut epithelium contains mainly SAα(2,3) moieties (*27*). The receptors with α(2,6)-linked sialic acid are often called "human-like receptors"; similarly, the receptors with α(2,3)-linked sialic acid are often called "avian-like receptors". The type of the receptors on the host cell surface determines the attachment efficiency

(leading to the infection) of influenza virus. In a recent study on the adaptation of an reassortant avian H5N1/human H1N1 virus in ferret (*28*), it shows that the HA of avian H5N1 which only binds to SAα(2,3)Gal shifts to bind SAα(2,6)Gal after mutations of two key amino acids in HA, demonstrating the adaptive ability of influenza viruses to fit the receptor requirement on the host cells for a better attachment. Indeed, the viral HA is usually known as a host-range determinant (*29*, *30*).

Following the attachment, the influenza virus is internalized mainly by clathrin-mediated endocytosis (*31*). Other internalization pathways include those through caveolae, through non-clathrin/ non-caveolae, and through macropinocytosis (*32*). The low pH environment in the endosome triggers the conformational change of HA, and activates the fusion of endosomal and viral membranes, releasing the RNPs into the cytoplasm of host cell. The M2 protein with ion channel activity allows the influx of protons into the virion, and the protons cause the dissociation of the RNPs from the other viral components, thus facilitating the uncoating of influenza virus (*33*). The nuclear localization signals (NLSs) are found on all the proteins of the RNP complex (PB1, PB2, PA, and NP) to interact with the nuclear import machinery for the import of viral RNA into the nucleus (*1*).

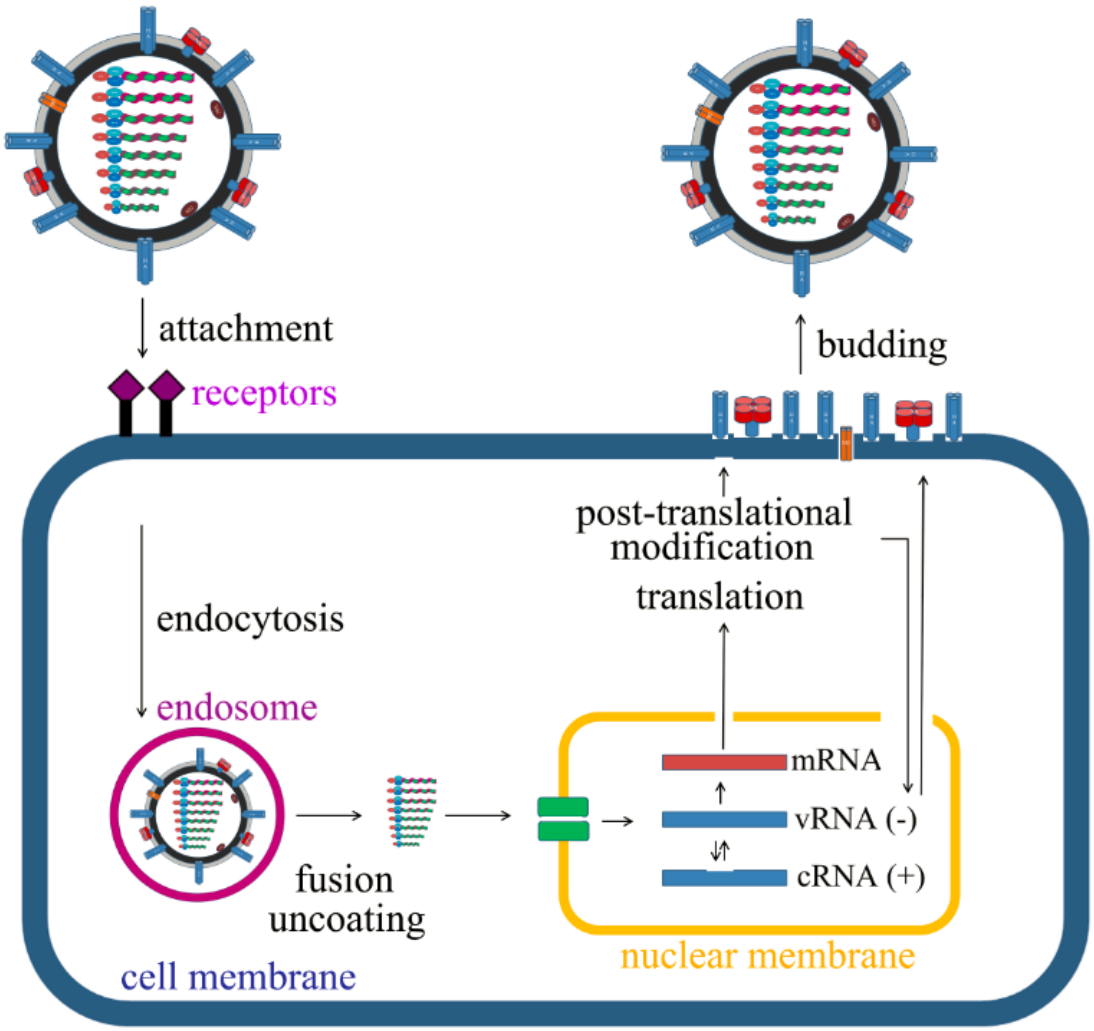

Figure 1.2 Life cycle of the influenza virus. The infection of the influenza virus begins with the attachment of virus particle to the recepotrs on the host cell surface. The bound virus is internalized through endocytosis. With the elevated pH in the endosome, the virus is fused with the endosomal membrane. The fusion and the uncoating of the virus result in the release the viral genome into the cytoplasm. The viral genome enters the nuclear through the nuclear import machinery. The vRNA is replicated by the viral polymerase complex, and the viral mRNA is also synthesized for the expression of viral proteins. The expressed viral polymerase and nucleoprotein associate with the vRNA to form new RNPs. The M1 and NS2 proteins bind with RNPs for their transportation. The NS1 protein is the main effector to suppress host immune response. The HA, NA, M2 proteins concentrate on the apical membrane for budding. The NA protein cleaves the sialic acid to release the progeny virions.

After entering the nucleus, the negative-sense viral RNA (vRNA) serves as the template to make a full-length positive-sense copy of complementary RNA (cRNA) and in turn serves as a template to make more vRNA, where different subunits of viral polymerase complex are involved in these steps (*34*). On the other hand, the vRNA is also transcribed into messenger RNA (mRNA) that is an incomplete products from vRNA template (*34*). The viral polymerase complex snatches a 5'-cap from host pre-mRNA transcripts to be added at the 5'-end of viral mRNA, and this process is catalyzed by the cap-binding function of PB2 (*35*) and endonuclease activity of PA (*36*). Thus, the cellular RNA polymerase II is utilized to initiate the synthesis of influenza virus mRNA.

The viral gene expressions are temporally regulated (*1*). At the early stages, the synthesis of NP and NS1 proteins is predominant possibly because of the particular roles of these proteins in the virus life cycle at the early stages (*37*). The NP protein is required for the assembly of newly replicated vRNA into RNP (*38*) while the NS1 protein is the important virulence factor of influenza virus to inhibit the host innate and adaptive immunity (*39*). The synthesis of HA, NA and M1 proteins is delayed (*37*). The HA and NA proteins are involved in the budding of progeny virions and the M1 protein together with NS2 dictates the nuclear export of RNP, and all these proteins are required at the later stages of virus life cycle (*40*).

The HA, NA, and M2 proteins concentrate at the apical membrane of polarized cells through their apical sorting signals and their interactions with lipid raft (*41*). The M1 protein serves as the recruiter to associate RNP with the

membrane at the budding site (*42*). The accumulation of the M1 protein also initiates the outward curvature of the apical plasma membrane, leading to the budding process (*41*). After the membrane fuses at the base of the bud, the enveloped virus still attaches to the host cell via the interaction of HA with the receptors on the cell surface. The NA protein is required to cleave the sialic acids in order to release the viral particles, and it also promotes the spread of the viral particles to infect the surrounding cells (*40*).

HA is a type I integral membrane protein (Figure 1.3), it is expressed as a precursor, called HA_0. HA_0 has a signal peptide at its N-terminus that directing its transportation to the cell surface. HA_0 can be cleaved to HA_1 and HA_2. HA_2 has a fusion peptide at it N-terminus. The transmembrane domain is located at the C-terminus of HA_2.

Figure 1.3 Linear sequence structure of hemagglutinin

HA mediates two important steps in the life cycle of influenza virus, binding and fusion; and it is also considered as one of the determinants for host switching (*29*, *30*). The studies on HA include the structural basis, the receptor binding specificity, the fusion ability, its interaction with antibodies, and the role in virus pathogenesis; all investigations are aimed to establish a more effective method to

fight against this lingering health threaten. As an effort to contribute more understandings on the influenza virus, influenza B virus HA (BHA) is chosen to deeply study its structure and function.

1.4. The anisotropic *B*-factor refinement

Structural characterization of the proteins of influenza virus provide the basic information for rational design of the anti-virus drugs and vaccines (*43*). Especially, the structure of HA in complex with antibody illustrates the interaction pattern utilized by antibody to recognize the HA antigen. By searching the protein data bank, it is found that many HA-antibody complex structures were determined at resolution below 3 Å (*44*–*47*). The low-resolution feature of these structures might prevent the accurate elucidation of the interactions in the model. As an attempt to improve the structure quality, the application of normal mode refinement (*48*) for the structures determined at the low resolution is tested.

For structural refinement in X-ray crystallography, temperature *B*-factors are used to take into account structural deviations from average positions of atoms. Conventionally, the atomic positional deviation is modeled as a spherical Gaussian, and a single scalar constant, the isotropic *B*-factor, is used to model the temperature factor for each atom independently. However, the large number of independent thermal parameters in the isotropic *B*-factor model, generally equal to the number of non-hydrogen atoms, imposes a severe burden for lower-resolution refinement due to low data-to-parameter ratio (risk of over-fitting). Moreover, the

independence among atomic thermal factors across proteins, in many cases, results in poor interpretation of collectiveness of molecular motions. This problem is a particularly severe for large and flexible complex structures where functions often involves long-range collective deformations. Finally, many large complexes contain highly flexible structural components that undergo orientation-specific (anisotropic) deformations. Ideally, anisotropic *B*-factors should be used for a more accurate description of those deformations. However, very frequently, the highly flexible components in large complexes compromise the resolution of diffraction, making impossible a full-scale anisotropic refinement that requires three positional and six thermal parameters for each atom for many such systems.

1.4.1. Normal mode refinement

Normal mode refinement (NMRef) provides a method to overcome the issues discussed above, because this approach can describe the collective motion with a small set of low-frequency modes. The idea of applying NMRef to X-ray crystallography was first proposed about 20 years ago, but successfully application to data with only moderate diffraction resolution data was achieved only recently, and was based on a new elastic network model for C_α-coarse-grained normal mode calculation to avoid the so-called "tip effect" (*49*).

In the new elastic network model, an extra term (Equation 1.1) was added to the conventional potential function with its minimum unchanged, while imposing a scaling factor on those flexible degrees of freedom. In this way, the new method could retain the advantage of conventional elastic normal mode analysis (eNMA)

that no initial energy minimization is needed; and meanwhile, this method suppressed the tip effect by attributing a large force constant to the flexible region.

$$V = (\gamma/2) \sum_i \sum_j h_{ij}(|\mathbf{r}_{ij}| - |\mathbf{r}_{ij}^0|)^2 + (\omega/2) \sum_\alpha (\phi_\alpha - \phi_\alpha^0)^2$$

$$h_{ij} = \begin{cases} 1 & |\mathbf{r}_{ij}^0| \le r_c \\ 0 & |\mathbf{r}_{ij}^0| > r_c \end{cases} \quad \omega = \zeta \min(H_{\alpha\alpha}^0)$$

Equation 1.1 Potential function for the new elastic network model

There are two important parameters to implement the new potential function as shown in formula 1.1. One is the cutoff distance (r_c) in the Heaviside step function h_{ij}, specifying the effective scope in which the coarse-grained points would be included in the calculation. The other one is scaling factor (ζ) in the extra term, representing the amount of stiffness that would be introduced to the angular terms. These two parameters could be adjusted to obtain more physically meaningful normal modes.

With normal modes in hand, the Debye-Waller factor could be expressed in Equation 1.2

$$\exp\left[-\frac{1}{2} \langle (\mathbf{q}^{\mathrm{T}} \Delta \mathbf{r}_j)^2 \rangle \right] \approx \exp -[\mathbf{q}^{\mathrm{T}} \mathbf{E}_j \langle \sigma \sigma^T \rangle \mathbf{E}_j^T \mathbf{q}]$$

Equation 1.2 Debye-Waller factor

Here, E_j is a $3 \times M$ eigenvector matrix for atom j with each column vector representing one normal mode, where M is the number of low-frequency modes. The variances of $\Pi_{mm} = <\sigma_m \sigma_m>$ represent the contributions from the effective normal modes, while the co-variances of $\Pi_{mn} = <\sigma_m \sigma_n>$ represent the contributions from the mode-mode coupling, causing the orthogonal rotation of the normal mode eigenvectors. To assure the positive semi-definiteness of matrix Π, it is expressed in terms of a lower triangular matrix as $\Pi = \Omega\Omega^T$ [2-9]. In the refinement, the elements of matrix Ω are treated as independent parameters for the minimization of the crystallographic residual. If there are M low-frequency modes used in the refinement, the number of the parameters accounting for the internal motion by normal mode is $M(M+1)/2$, which would be much smaller than that used for conventional isotropic B-factor refinement (*50*).

1.4.2. Translation, libration and screw-rotation (TLS) refinement

Whereas normal modes could describe the internal motion of the macromolecule in its biological aspects, TLS represents the external motion of the

pseudo-rigid bodies within the crystal (*51*). The bodies could be the whole molecule, functional domains, or secondary elements. The movement of a rigid body is described as the combination of a rotation about an axis passing through the origin and a translation of the origin (*52*).

The displacement of an atom in a rigid body can be expressed by a set of 20 parameters, constituting three *3×3* tensors: symmetrical tensors **T** and **L** describing the translational and rotational components of the rigid-body displacement and generally non-symmetric tensor **S** describing the correlation between translational and rotational displacements (*53*, *54*). The number of the total parameters used in the refinement depends on how many rigid bodies are defined in the asymmetric unit.

Despite its widely observed success, TLS refinement has certain limitations, most of which are related to its rigid-body assumption. For example, the motions between rigid-bodies are not interdependent, thus making the TLS model less collective for large-conformational deformations. Another issue of TLS is that the grouping scheme in multi-group refinement can be tricky. The best systematic way of grouping is derived from the distribution of converged isotropic *B*-factors. Such a feature raises the issue of reliability, or even the possibility, of successful grouping in cases where the isotropic *B*-factors are poorly, or not at all, converged — a case often seen in structural refinement at moderate resolutions.

1.4.3. Selection of the structures

A set of systems determined at the low resolution are tested to demonstrate when and how NMRef is employed to improve structural refinement in X-ray crystallography. In the eight studied structures, five of them are complex structures, while the other three are determined at 3.5 A to 3.9 Å.

The first system is the structure of the voltage-dependent potassium channel (KvAP) in complex with monoclonal Fv fragment at 3.9 Å (PDB accession code: 2A0L) (*55*). The voltage-dependent potassium channel forms the potassium selective pores across the membrane to allow the flow of potassium in responsible to the membrane potential, and the monoclonal Fv fragment was included to improve the crystal resolution (*55*).

The second one is the structure of the CorA magnesium transporter at 3.9 Å (PDB accession code: 2BBJ) , which is responsible for the uptake of magnesium in prokaryotes (*56*). Several principles of transportation of magnesium through the CorA magnesium transporter were proposed based on the structure analysis (*56*).

The third one is the complex structure of yeast Sec13/31 at 3.3 Å (PDB accession code: 2PM9), which is one of the assembly units of COPII coat (*57*). By fitting this crystal structure into an electron microscopy density map of COPII cage, a molecular model of the COPII cage was constructed (*57*).

The fourth one is the structure of HIV-1 reverse transcriptase (RT) in complex with Janssen-R185545 (an anti-AIDS drug candidate) at 3.0 Å (PDB

accession code: 1SUQ), that reveals the conformation of drug that adopted to fit into the hydrophobic binding pocket on HIV-1 RT (*58*).

The fifth one is the structure of an open state of the mechanosensitive channel (MscS) at 3.45 Å (PDB accession: 2VV5), that shows a pore with 13 Å in diameter (*59*). The mechanism that underlies the gating of the MscS was suggested based the structure (*59*).

The sixth one is the structure of 1998 human influenza virus neuraminidase in complex with its antibody at 3.0 Å (PDB accession: 2AEQ) (*60*). The structure illustrates a novel interaction pattern as compared to other NA-antibody structures, and it supports the concept that antibody selection of NA escaping variants is important for the antigenic drift in human H3N2 influenza viruses (*60*).

The seventh one is the structure of chicken embryo lethal orphan (CELO) adenovirus major coat protein hexon at 3.9 Å (PDB accession code: 2INY) (*61*). A remarkable channel with 100 Å in diameter that runs through the entire hexon crystal reveals its potential application as nano-material (*61*).

The last system is the complex structure of *Deinococcus radiodurans* RecR and RecO (drRecOR) at 3.8 Å (PDB accession code: 2V1C) (*62*). The complex shows a RecR:RecO molecular ratio of 2:1, in consistence with the previous biochemical studies (*62*). The structural analysis together with the mutagenesis studies indicates a mechanism that underlies the recognition of stalled DNA replication forks by drRecOR (*62*).

All these systems are featured as low data-to-parameter ratio, un-converged *B*-factor profile, which are possible to happen to the crystal structures of HA-antibody complex. The studies on these systems with normal mode refinement should provide some useful guidelines for the further refinement of HA-antibody structures.

1.5. Receptor binding ability of BHA

The ability of influenza A virus to efficiently infect human population is correlated with the high affinity of HA to α(2,6)-linked cell-surface sialic acid receptors with a characteristic umbrella-like structural topology (*63*, *64*). The receptors bind to a shallow receptor-binding site (RBS) in the globular domain of HA formed by three key residues that are absolutely conserved among all known influenza A virus HA (AHA) glycoproteins, Tyr98, Trp153, and His183 (Figure 1.4) (H3 AHA numbering) (*65*).

Different from influenza A virus, influenza B virus does not have subtypes. Instead, currently circulating influenza B viruses belong to either Yamagata lineage or Victoria lineage that diverged from each other in the early 1970s (*66*, *67*). Most strikingly, as mentioned above, influenza B virus has a very limited host range, circulating almost exclusively among humans (*9*). In measuring the binding affinity of influenza B virus to receptor analogs *in vitro*, it was found that influenza B viruses overall have lower binding affinities as compared to those of influenza A virus (*68*).

Our group has determined the crystal structures of influenza B virus HA (BHA) in various states (*69*, *70*). The overall structure of BHA is similar with that of AHA (*69*). Comparison of the structures of influenza AHA and BHA glycoproteins revealed a critical difference in the receptor binding sites: Tyr98 in AHA *versus* Phe95 in BHA (based on the numbering of influenza B/Hong Kong/8/73 HA) (*69*, *70*). The hydroxyl group of Tyr98 in AHA contributes two hydrogen bonds to the bound Sia-1 moiety, which are lost in BHA when Phe95 is present instead (Figure 1.4B) (*70*). It is hypothesized that the loss of the hydroxyl group at Phe95 in BHA might contribute to the low binding affinity of influenza B virus (*71*).

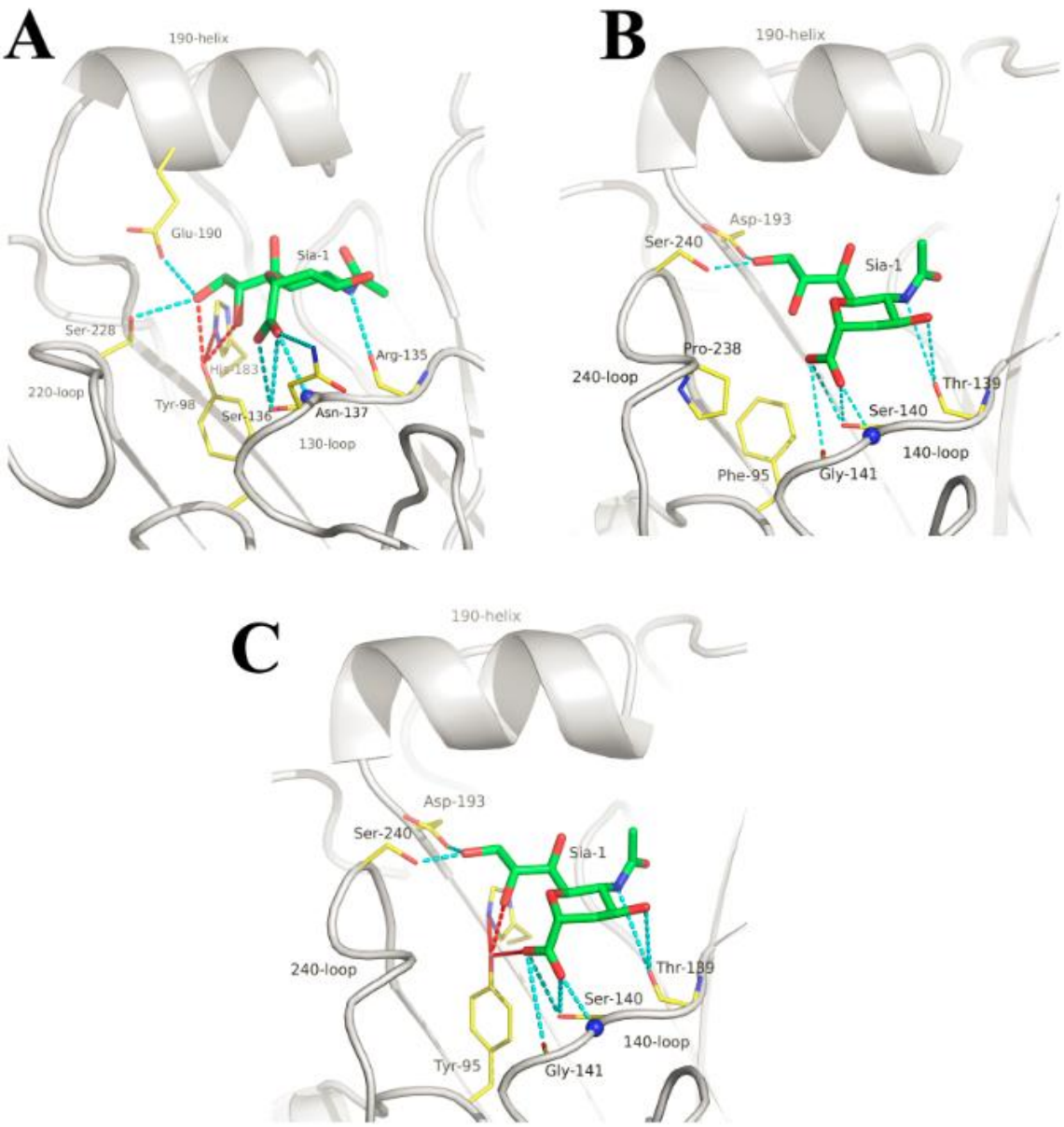

Figure 1.4 Comparison of the receptor-binding sites of influenza A and B virus HA. Shown are (A) influenza A/Hong Kong/68 H3 HA (PDB accession code: 1HGE); (B) influenza B/Hong Kong/8/73 HA (PDB accession code: 2RFU); (C) influenza B/Hong Kong/8/73 HA with the mutation Phe95→Tyr modeled *in silico*.

In order to test this hypothesis, the Phe95→Tyr mutation was generated by using site-directed mutagenesis. Although naturally occurring influenza B viruses generally have asparagine (Asn) and serine (Ser) at HA_1 194 and 196 respectively, creating a glycosylation site at Asn194, this site was frequently abolished in egg-adapted variants (*72*–*78*) and sporadic field isolates (*79*, *80*), as a result of mutations at HA_1 194 or 196 or both. Thus, the mutation Phe95→Tyr was constructed on BHA with (wild type) or without (Asn194→Asp) a glycosylation site at HA_1 194. Recombinant BHA proteins are used to demonstrate the critical roles of Phe95 in determining the binding affinity for a variety of sialic acid receptors and cultured cells. Recombinant viruses are also generated by reverse genetics to characterize the role of Phe95 in receptor binding ability at virus level.

1.6. Membrane fusion induced by HA

Membrane fusion that results in the mixing of enveloped genomic materials with the contents of the host cell is an important step in viral life cycle (*81*). Viruses utilize their surface proteins, usually called fusion proteins, to overcome the energy barrier required to merge the viral and host membranes (*82*). Though the sequences and structures of fusion proteins from different families of viruses are extremely diversified, recent studies suggest a common mechanism by which fusion proteins proceed from initial pre-fusion states, through a six-helix bundle intermediate state, to the final trimeric hairpin structure. The conformational

changes of fusion proteins can be triggered by either low pH, receptor binding or other novel factors (*81*, *83*, *84*).

Among the many characterized fusion proteins, hemagglutinin of influenza A virus (AHA) serves as a paradigm to study the protein-mediated membrane fusion (Figure 1.5) (*85*). The crystal structures for fusion-incompetent, pre-fusion and post-fusion states of AHA are well established by several groundbreaking papers (*65*, *86*–*88*). The structure of the precursor of AHA (AHA_0) shows that the C-terminus of HA_1 and N-terminus of HA_2 form a surface loop that is susceptible for cleavage by trypsin-like enzymes (Figure 1.5A), and HA_1 acts like a clamp sitting atop the stem region formed by HA_2 (*86*). Cleavage of HA_0 results in HA_1 and HA_2, which are linked together by one disulfide bond (Figure 1.5B). The fusion peptide at the N-terminus of HA_2 inserts into a hydrophobic cavity near the bottom of the stalk (*65*).

When pH decreases, HA_1 dissociates from HA_2 due to the protonation of multiple ionizable residues locating at the interface between HA_1 and HA_2 (Figure 1.5C). The loop-to-helix transition of the B-loop of HA_2 in its pre-fusion state elongates the coiled-coil structure in the pre-fusion state, by the newly formed coiled-coil from helix B and helix A. The elongated coiled-coil projects the fusion peptide into the target membrane (Figure 1.5D). The loop-to-helix transition of the B-loop is central to the so-called “spring-loaded mechanism” that reflect the large free energies released from this transition and are utilized to overcome the energy barrier of membrane fusion (*89*–*91*).

A small part of helix C rearranges (Figure 1.5E) so that the hydrophobic interactions in this region are altered (*87*, *88*). The rearrangement results in the formation of a six-helix bundle(Figure 1.5F), as accompanied by the curvature of both viral and host membranes. The complete formation of the six-helix bundle is important for the fusion intermediate hemi-fusion, where only the inner lipid leaflets merge while the outer lipid leaflets remain intact (Figure 1.5G). The HA_2 protein continues to collapse due to the interactions between the C-terminal segment and the groove formed by the central helices, until the fusion peptide and the transmembrane domain of HA are in close proximity (Figure 1.5H). Multiple HA_2 around the fusion site work together to open and enlarge the fusion pore, resulting in the mixture of viral and cell contents.

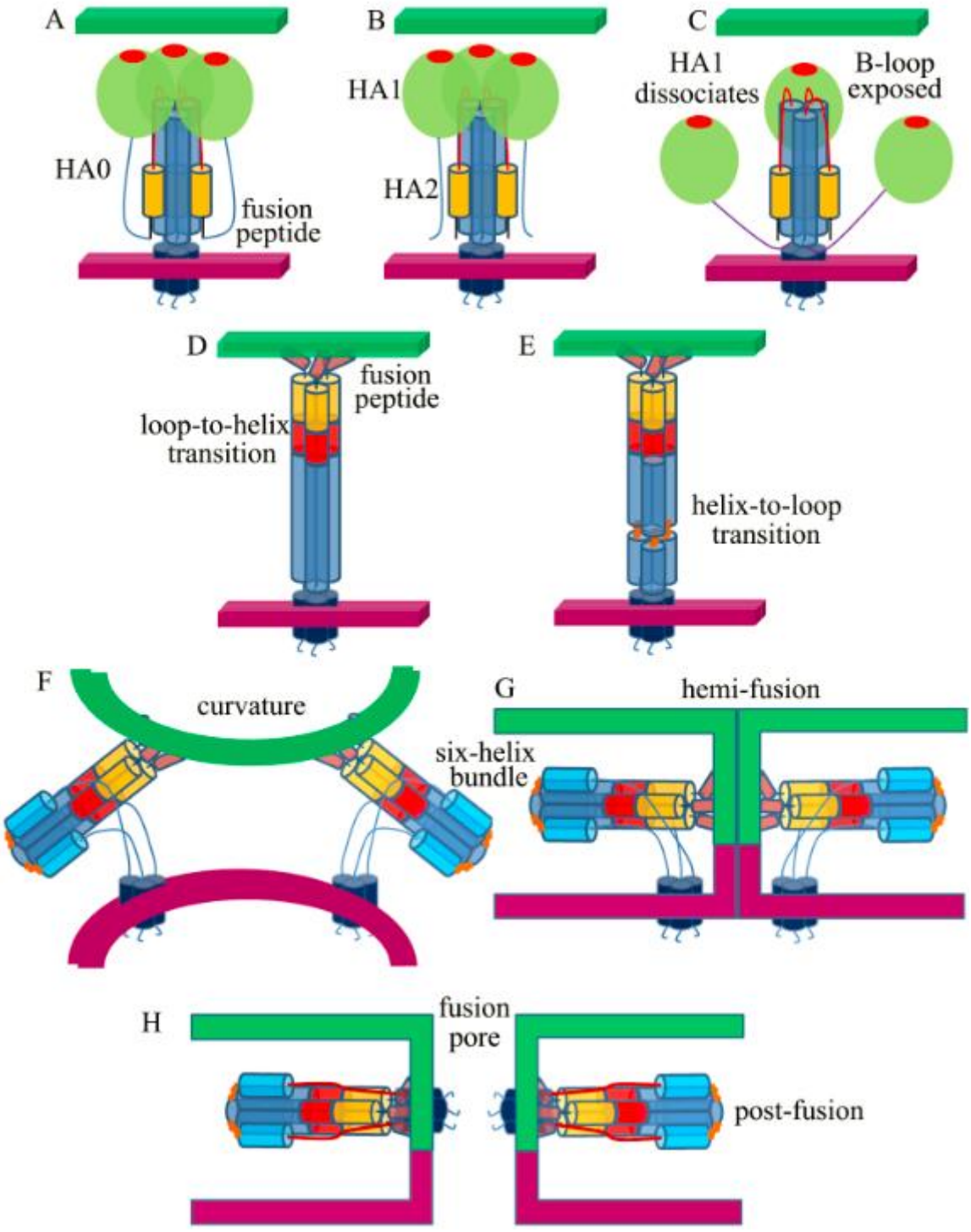

Figure 1.5 Model for the fusion process induced by HA.

Influenza B virus may cause other severe diseases, including myositis and leukopenia in children and is responsible for seasonal influenza epidemics every 2-4 years (*92*–*95*). As mentioned above, the structure of BHA in its pre-fusion state has only been determined recently and has provided much information on its antigenetic sites and receptor binding specificities (*69*, *70*). Hemagglutinin of influenza B virus (BHA) shares only 28% protein sequence identities with AHA (*96*). The mechanism underlying the fusion process induced by BHA is largely unknown without structural information of BHA at post-fusion state.

Chapter 2

Methods and Materials

2.1. Anisotropic *B*-factor refinement

2.1.1. Normal mode refinement

In order to eliminate the differences in *R*-factors from different refinement programs, the original model was first used as input to REFMAC5 (*97*) to update the *R*-factors, which were used in future comparison with normal-mode-refined structures. In the normal-mode-based refinement, optimal values for cutoff distance and stiffness were empirically determined for each individual system, as judged by the lowest R_{cryst}- and R_{free}-values in subsequent normal-mode refinement. A different number of normal modes (M), corresponding to a range of different refinement parameters (equals to M(M+1)/2 + 27), were used in normal-mode refinement to generate anisotropic *B*-factors. The anisotropic *B*-factors then replaced the isotropic *B*-factors in the original model, which was then subject to

REFMAC5 (*97*) refinement with very tight geometric restraints for positional refinement. The R_{cryst}- and R_{free}-factors were monitored throughout, following the Equation 2.1:

$$R = \frac{\sum_{\mathbf{h}} \left\| F_{obs}(\mathbf{h}) \right| - \left| F_{cal}(\mathbf{h}) \right\|}{\sum_{\mathbf{h}} \left| F_{obs}(\mathbf{h}) \right|}$$

Equation 2.1 *R*-factor

where the same set of free reflections used in the original structure determination was saved for calculation of the R_{free}-factor.

To compare with the isotropic *B*-factor profiles in the original model, the anisotropic *B*-factor profile of the normal-mode model was converted to isotropic ones by averaging the diagonal terms of anisotropic displacement parameters for each atom.

2.1.2. TLS refinement on NMRef structures

The normal-mode refined structure was submitted to the TLS Motion Determination (TLSMD) web server (*51*). Models with different partition of each peptide chain would be generated. Each model was described by a *tlsin* file containing the parameters fitted for each TLS group and a *xyzin* file containing the atomic coordinates from the input model with modified B_{iso} values. These two files were used in REFMAC5 (*97*) to carry out five cycles of TLS-parameter refinement

followed by five cycles of restrained positional refinement. The TLS-parameter refinement and positional refinement were repeated until a lowest R_{free}-value was achieved.

2.2. Functional study of Phe95→Tyr mutation of BHA

2.2.1. Cells

BHK21 (Baby hamster kidney cell), Vero (African green monkey kidney cell), 293T (Human kidney cell), CV-1 (Green monkey kidney cell) and A549(Human lung cell) cells were maintained in DMEM (Dulbecco's Modified Eagle Medium) with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin in a CO_2 incubator. MDCK cells were maintained in EMEM (Eagle's Minimal Essential Medium) with 10% FBS and 1% penicillin-streptomycin.

2.2.2. Cloning, expression and purification of BHA and its mutants

The cDNA encoding HA of influenza B/YM/73 (BHA) virus (a kind gift from Dr. Peter Palese) was cloned to pRB21 plasmid (a kind gift from Dr. Bernard Moss) to allow the expression in vaccinia systems (*98*). Site-directed mutagenesis was used to introduce the Phe95→Tyr, Asn194→Asp and Phe95→Tyr/Asn194→Asp mutations into BHA (Agilent Technologies). All the mutant constructs were fully sequenced. The recombinant vaccinia viruses were generated by infecting CV-1 cells with wild type virus vRB12 and then transfected with pRB21 vector containing BHA or its mutants (*98*). All the constructs were expressed as the unprocessed HA0

containing a C-terminal trimerizing sequence from the bacteriophage T4 fibritin (foldon) to facilitate the trimerization and expression of HA in the absence of the transmembrane region (*99*) and contained a His-tag at the extreme C-terminus. The CV-1 cells were maintained in DMEM medium supplemented with 10% fetal bovine serum (FBS). For a typical preparation, 400 mL confluent CV-1 monolayer cells were infected by recombinant vaccinia viruses at a multiplicity of infection (MOI) of 1. After three days, the supernatant was collected and dialyzed against buffer of 20 mM Tris-HCl (pH 7.2), 50 mM NaCl. The soluble BHA was recovered from the supernatant by HisPur Cobalt Resin (Thermo Fisher Scientific Inc.) The beads bound with BHA were treated with neuraminidase (Sigma) to prevent sialic acid-mediated aggregation. The BHA was eluted with 500 mM imidazole and was exchanged to buffer of 20 mM Tris-HCl (pH 7.2), 50 mM NaCl by three cycles of concentration and dilution with a 30 kD-cutoff centrifugal filter unit (Amicon), then subjected to anion-exchange (mono-Q 4.6/100 PE, GE Healthcare) and gel-filtration chromatography (Superdex 200 10/300 GL, GE Healthcare). The peak corresponding to the BHA-trimer was collected and concentrated. Protein concentration was quantified by Bradford protein assay (*100*).

2.2.3. Dose-dependent glycan assay

To characterize the receptor binding properties of BHA and its mutants, biotinylated α(2, 3)-glycan, 3'SLN-LN-LN, and α(2, 6)-glycan, 6'SLN-LN, from the Consortium of Functional Glycomics (www.functionalglycomics.org) were used in the dose-dependent glycan assay (*64*, *101*). LN represents lactosamine (Galβ1-

4GlcNAc), 3'SLN and 6'SLN represent Neu5Acα2-3 and Neu5Acα2-6 linked to LN, respectively. The different glycans were first diluted to 2.4 μM and loaded on to the streptavidin-coated high binding capacity 384-well plates (Pierce), followed by incubation at 4 degree overnight. Excessive glycans were removed by washing with PBS three times. The pre-complex of BHA, mouse anti-His antibody (0.2 mg/mL, Sigma) and anti-mouse-IgG antibody (2 mg/mL, Sigma) was prepared in a molar ratio of 4:2:1. The mixture was incubated on ice for 20 minutes, and the pre-complex was diluted to the expected concentration with 1% BSA in PBS. Each glycan-coated well was then loaded with 50 μL of the pre-complex of BHA and incubated at room temperature for 2 hours followed by wash with 0.05% Tween-20 in PBS to remove any unbound pre-complex. The Amplex Red Peroxidase Assay kit (Invitrogen) was used to measure the binding signal based on the HRP activity. Negative controls, where no pre-complexes were loaded in those wells, were included for each glycan, and the assays were performed in triplicate. To quantify the binding affinity, the assays were performed at low BHA concentration range (0.05 to 1.0 μg/mL) for 3'SLN-LN and 6'SLN-LN. A linearized Hill equation was used to represent the binding of the pre-complex of BHA to the coated glycan (with the data from 0.05 to 1.0 μg/mL) (*64*, *102*).

2.2.4. Glycan microarray analysis

The binding profile of purified influenza B virus hemagglutinin and its mutants to different glycans was determined by the Consortium of Functional Glycomics (www.functionalglycomics.org) using version 4.1 glycan array, which

contains 465 glycan structures in replicates of six. The BHA proteins were first incubated with anti-penta His antibody (Qiagen) and Alexa488-labeled anti-mouse IgG (Invitrogen) on ice for 30 minutes. The final concentration of BHA proteins used in the assay was 200 μg/ml. The protein mixture was then applied to a fresh hydrated glycan array under a cover slip. The fluorescence signals were recorded. This assay is carried out by the Consortium of Functional Glycomics (www.functionalglycomics.org). For those glycans with strong signals indicating good binding, the relative fluorescence unit (RFU) value are about 10,000 to 50,000. To make the data clear, an arbitrary value of 5000 RFU is set to be the cutoff for all protein samples. If the glycan shows signal below 5000 RFU, it is regarded as a weak or un-recognized glycan. The glycans that did not show signals above 5000 RFU for any of BHA or its mutants are not considered. Most of those un-shown glycans are without sialic acids.

2.2.5. Red blood cell binding assay

Thirty microliter Ni-NTA resin (Thermo Fisher Scientific Inc.) was loaded with 2.5 μg BHA and incubated at room temperature for 2 hours. The beads were centrifuged, and the unbound BHA in supernatant was removed. The beads were then blocked with 3% BSA in PBS for 30 minutes. The supernatant was removed after spinning down the beads, and the beads were washed with PBS extensively. The red blood cells (RBCs) were washed with Alsever's solution until the supernatant is clear. The RBCs were stored as 50% in Alsever's solution after the final wash. The RBCs were then added to the BHA-coated beads to a final

concentration of 0.5% and incubated at room temperature for 30 minutes. Finally, the beads were washed gently with PBS for appropriate imaging under microscope.

2.2.6. Influenza A virus infection inhibition assay in MDCK cells

For this assay, the BHA eluted from the Cobalt Resin was used. The purity of BHA was about 85% to 90% as determined by SDS-PAGE. MDCK cells were seeded at appropriate concentration in a 96-well plate for 24 hours to achieve 100% confluency for performing the assays . The cell monolayer was washed with Dulbecco's phosphate-buffered saline (DPBS) twice and complete Dulbecco's modified eagle medium (DMEM) (0.2% BSA, 25 mM HEPES in DMEM) once. The expected amount of BHA was loaded into different wells and complete DMEM was added to make all the wells contain same volume. After the plate was incubated at room temperature for 1 hour, 5 HAU (Hemagglutination units) per 50 μl 1% human RBC of influenza A/Brisbane virus and TPCK-trypsin (final concentration was 2 μg/ml, Worthington) were added to wells, and the plate was incubated at 37 degree for 1 hour (5% CO_2). The supernatant was removed, and the wells were washed twice with complete DMEM. Finally, each well was loaded with 100 μl influenza virus growth medium (complete DMEM with 2 μg/ml TPCK-trypsin). After 24 hours of incubation, the supernatant was centrifuged (300 g, 15 minute) to pellet the cellular debris and used for the standard hemagglutination assay with 1% human RBC. All assays were performed in triplicate.

2.2.7. Cell-based enzyme-linked immunosorbent assay (ELISA) assay

In order to quantitatively detect the binding of receptors on the cell surface to BHA, cell-based ELISA assay was carried out with three cell lines including MDCK, BHK21 and Vero in 96-well plates. For this assay, the BHA eluted from the Cobalt resin was used. The confluent cell monolayer in white 96-well plate (FALCON) was first washed with DPBS and different amount of BHA (0.1, 0.2, 0.5, 1, 2, 5, 10, 15, 20, 30 μg/100μl) diluted with 1% BSA in DPBS was added to different wells. Complete DMEM was then added to take all wells to the same volume of solution. The plate was incubated at room temperature for 1 hour. The supernatant was removed and each well was washed with DPBS for three times followed by addition of 100 μl 3.7% formaldehyde diluted in DPBS for 10 minutes. The fixative was removed, and wells were washed with PBST (0.1% Triton X-100 in PBS [Phosphate buffered saline]) for three times. The wells were then blocked with 10% FBS for 1 hour at room temperature followed by incubation with primary antibody (mouse anti-His antibody, Sigma, 1:4000 dilution with 1% BSA in PBST) for 1 hour at room temperature. The wells were then washed three times with PBST. As a final step, secondary antibody (HRP-conjugated anti-mouse-IgG, Sigma, dilution 1:40000) was added to the wells for 1 hour at room temperature. After washing with PBST and PBS, signal from each well was measured with ECL reagent (Amersham). All assays were performed in triplicate.

2.2.8. Reverse genetics and virus stock preparation

The plasmids for reverse genetics of influenza B/Lee/40 virus were generous gifts from Dr. Kawaoka (Department of Pathobiological Sciences, University of Wisconsin-Madison). The protocol to generate recombinant influenza viruses was generally according to a previous report (*103*). The mutations Phe95→Tyr, Asn194→Asp and Phe95→Tyr/Asn194→Asp were introduced to BHA by site mutation. Co-cultures of 293T and MDCK in DMEM with 10% FBS were plated one day before. Two microliter TransIT-LT1 (Panvera, Madison, WI) was used per one micro-gram plasmid. The preparation of DNA-lipid complex was according to the manufacturer's instructions. The DNA-lipid mixture was added into the 293T-MDCK co-culture, and the plate was incubated at 33 degree. After 16 hours, the medium was removed, and the cell monolayer was washed twice with OMEM. Fresh OMEM with 1 μg/ml trypsin was added, and the plate was further incubated for 48 hours. For virus with Phe95→Tyr mutation in HA, 1 μg/ml neuraminidase (Sigma) was added to compensate for the increased binding ability of this variant. The supernatant was harvested and was then propagated on MDCK cells. The supernatant containing viruses was then plaque-purified and was propagated twice in MDCK cells. MDCK monolayer in T-75 flask after overnight incubation was washed with plain EMEM for three times, followed by infection of 1 ml virus at room temperature with occasionally gentle rocking. After 1 hour, the inoculum was removed and 30 ml EMEM with 1 μg/ml TPCK-trypsin was added. The cells were incubated at 33 degree for 3 days until at least 70% cells shown cytopathic effect

(CPE). The supernatant was harvested, and clarified at 2600 g for 5 min, and layered over a 25% sucrose cushion in NTE buffer (100 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl, pH 7.4). The virus was concentrated by ultracentrifugation at 30,000 rpm for 3 hours at 4 degree. The virus pellet was re-suspended in NTE buffer, aliquoted, and stored at -80 degree. The purified viruses were thawed before using, and the titer was determined with standard techniques.

2.2.9. Sequencing of virus stock

The virus stock was sequenced to confirm the desired mutation in HA gene. QIAamp viral RNA kit was used to extract the virus RNA. The HA gene was amplified with designed primers by QIAGEN Onestep RT-PCR kit. The sequenced region included nucleotides from about 150 to 1600. The viruses with Phe95→Tyr, Asn194→Asp and Phe95→Tyr/Asn194→Asp in HA gene have correct sequences. For the wild type virus, which was expected to have an Asn at site 194, it had the correct sequence after one passage in MDCK cells. But the glycosylation site at HA1 194~196 was quickly lost after the second passage in MDCK cells (Asn194 was mutated to Asp194 or Ser194 in five and one, out of six sequenced viruses, respectively). Because Asn194 was not stable in MDCK cells, only viruses have Asn194→Asp or Phe95→Tyr/Asn194→Asp in HA were used in the following studies.

2.2.10. Replication of recombinant viruses in MDCK and mice

For replication of recombinant viruses in MDCK, the cell monolayer was infected with viruses at MOI of 0.1. The supernatant was harvested after 48 hours infection. For the replication in mice, mice were administered virus intranasally at 10^4 PFU, and lungs were harvested after 3 days of infection (This experiment was performed by Dr. Lin Liu at Dr. Innocent Nnadi Mbawuike's lab, Department: Molecular Virology & Microbiology, Baylor College of Medicine). The titers are determined by plaque assay in MDCK. For plaque assay with Phe95→Tyr/Asn194→Asp virus, the plaque forms rod-like shape. To get circle-shape plaques, 1 μg/ml neuraminidase (Sigma) was added.

2.2.11. Binding assay of recombinant viruses

The binding of fluorescently labeled viruses to different cell lines was according to methods as reported (*104*). First, virus (50 μl) was incubated with Alexa 488 (25 μg, Invitrogen) in the presence of 0.1 M $NaHCO_3$ (pH 9.0) for 1 hour at 37 degree. Second, the excessive Alexa 488 was removed from the labeled virus by dialysis against PBS with 1 mM EDTA overnight at 4 degree. Slide-A-Lyzer MINI Dialysis Devices (7K MWCO, Thermo Scientific) was used for the dialysis. The labeled viruses were then used for the binding experiments on the next day. Third, confluent BHK21, Vero, A549 and MDCK cells prepared in 96-well plate one day before were chilled at 4 degree for 1 hour. The cell monolayer was overlaid with labeled viruses at a multiplicity of infection of 3 for 1 hour at 4 degree, washed with PBS for three times, and read by FLUOstar Omega (BMG LABTECH) using

excitation/emission at 485 nm/528 nm. The binding assays were performed in triplicate. The substrate MUNANA (2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid) was used to assay for viral NA activity at 4 degree.

2.3. Structure determination of BHA_2 at the post-fusion state

The protein HA_2 of B/HK/73 with His-tag at its N-terminus was expressed in *E. coli* and successfully produced high quality diffraction data in our lab. BHA_2 was concentrated to 10 mg/ml, crystals were grown in a 290 K incubation room by hanging drop using a 1:1 ratio of protein solution to reservoir solution of 0.1 M ammonium citrate, 13.5% PEG3350 (pH 7.0). The diffraction data were collected with LS-CAT beamlines at the Advanced Photon Source at the U.S. Department of Energy's Argonne National Laboratory (Chicago, IL). The diffraction images were indexed and integrated by HKL2000 (*105*) or Mosflm (*106*, *107*), followed by scaling using SCALA executed in the CCP4 suite (*108*). The intensities were then truncated to structure factor amplitude by TRUNCATE in CCP4 (*108*). Five percent of unique reflections were randomly chosen as test set for calculating the Free R value.

The long helix of AHA_2 (PDB accession code: 1QU1) including residues 37 to 105 was pruned to Cβ atoms by CHAINSAW in the CCP4 suite (*108*) and served as the search model for molecular replacement. One copy of the model was found by AutoMR module implemented in Phenix (*109*). The resulting σ-weighted 2FOFC map showed the densities for the other part of protein. The residues were built into the density by Coot (*110*)or O (*111*). The model was refined by Refmac5 (*97*)or

Refine module in Phenix. The Figures for structural snapshots were generated by Pymol (*112*).

Chapter 3

Results

3.1. Normal mode refinement improves the structure

3.1.1. Structure of KvAP-33H1 Fv complex at 3.9 Å resolution

The diffraction data for the KvAP-33H1 Fv complex contained 16,234 unique reflections in the range of 57.89~3.90 Å. The reported structural model (PDB accession code: 2A0L) refined using CNS (*113*) had 6,740 protein atoms and four potassium ions, with R_{cryst}- and R_{free}-factors of 35.8% and 39.2%, respectively (*55*). This model has an average *B*-factor of 188 Å^2, with many atoms reaching over 200 Å^2, indicating an overall high structural flexibility (Figure 3.1a, light line). Moreover, the original structure employed a *B*-factor cutoff of ~240 Å^2, making the *B*-factors for some highly mobile atoms underdetermined.

In order to eliminate the differences in *R*-factors that stemmed from using different refinement programs, the normal-mode refinement generally began with a re-calculation of the *R*-factors using the original model as input to REFMAC5 (*97*). The recalculated R_{cryst}- and R_{free}-factors for the KvAP-33H1 Fv complex were 34.9% and 37.9%, respectively. These values were used as reference when comparing with the normal-mode refined model. In order to capture the molecular deformations that are intrinsic to this molecule, the tetrameric biological unit was first generated from symmetry operations. The normal-mode calculation was performed on the tetrameric biological unit with the best combination of cutoff and stiffness values of 13 Å and 30, respectively. Only the portion of eigenvectors corresponding to the asymmetric unit, which was a half of the tetramer, were used in the normal-mode anisotropic *B*-factor refinement followed by positional refinement using REFMAC5 (*97*). The refinement yielded the lowest R_{cryst}- and R_{free}-factors of 30.0% and 35.0% with the inclusion of 48 modes (1203 parameters) (Figure 3.1b; Table 3.1), representing decreases of 4.9% in R_{cryst}-factor and 2.9% in R_{free}-factor. It is worth noting that the number of thermal parameters used in the normal-mode-based anisotropic *B*-factor refinement is more than five times fewer than in the original isotropic *B*-factor refinement.

In geometry analysis by MolProbity (*114*), the residues in the original model had a distribution of 77.3% in the Ramachandran favored region and 4.66% in the Ramachandran outlier region. After normal-mode refinement, the distribution became 84.5% and 3.03%, respectively, indicating a substantial improvement in geometry (Table 3.2).

Table 3.1 Statistics of R_{cryst}- and R_{free}-factors after normal-mode refinement and subsequent multi-group TLS refinement for all structures

PDB accession code	2A0L	2BBJ	2PM9	1SUQ
Initial R_{cryst}/R_{free}	34.9%/37.9%	34.8%/39.5%	23.9%/28.7%	26.0%/31.6%
NMRef-R_{cryst}/R_{free} (number of parameters)	30.0%/35.0% (1203)	29.6%/35.8% (1980)	22.4%/25.9% (42)	24.0%/29.0% (588)
NMRef+TLS-R_{cryst}/R_{free} (number of parameters)	35.1%/38.2% (2260)	29.3%/34.6% (1500)	21.1%/24.3% (600)	25.5%/29.9% (200)
PDB accession code	2VV5	2AEQ	2INY	2V1C
Initial R_{cryst}/R_{free}	29.3%/31.2%	28.5%/29.1%	34.4%/43.3%	43.1%/45.7%
NMRef-R_{cryst}/R_{free} (number of parameters)	26.1%/26.8% (147)	25.0%/27.4% (588)	31.4%/40.1% (105)	28.3%/38.0% (63)
NMRef+TLS-R_{cryst}/R_{free} (number of parameters)	25.8%/26.8% (420)	23.0%/27.0% (900)	32.3%/40.8% (200)	29.1%/37.9% (600)

Table 3.2 Comparison of residue distribution in the Ramachandran plot for the original models, normal-mode refined models and subsequent multi-group TLS refinement for all structures

PDB accession code	model	Ramachandran Favored (%)	Ramachandran Outlier (%)
2A0L	Original	77.3	4.66
	NMRef	84.5	3.03
2BBJ	Original	90.0	2.30
	NMRef	87.1	2.79
2PM9	Original	74.7	8.38
	NMRef	83.7	6.25
	NMRef + TLS	83.8	5.95
1SUQ	Original	78.9	4.31
	NMRef	85.5	2.77
2VV5	Original	92.4	2.75
	NMRef	93.8	2.35
	NMRef + TLS	94.0	2.35
2AEQ	Original	85.9	3.64
	NMRef	88.4	2.64
	NMRef + TLS	89.9	1.82
2INY	Original	59.2	14.5
	NMRef	65.5	8.87
2V1C	Original	90.8	2.26
	NMRef	77.1	5.33

The anisotropic *B*-factors of the normal-mode model were converted to isotropic ones and shown in Figure 3.1a as a dark line, which appeared to be a more realistic representation of the thermal motions of a protein at a resolution of 3.9 Å. When the ellipsoid of each C_α-atom was mapped to the structure (Figure 3.1c and Figure 3.1d), a higher structural flexibility was suggested for the atoms at the peripheral region of the structure.

However, the multi-group TLS refinement using the isotropic-equivalent *B*-factor profile from the normal-mode refined structure did not further improve the R_{cryst}- and R_{free}-factors (Table 3.1).

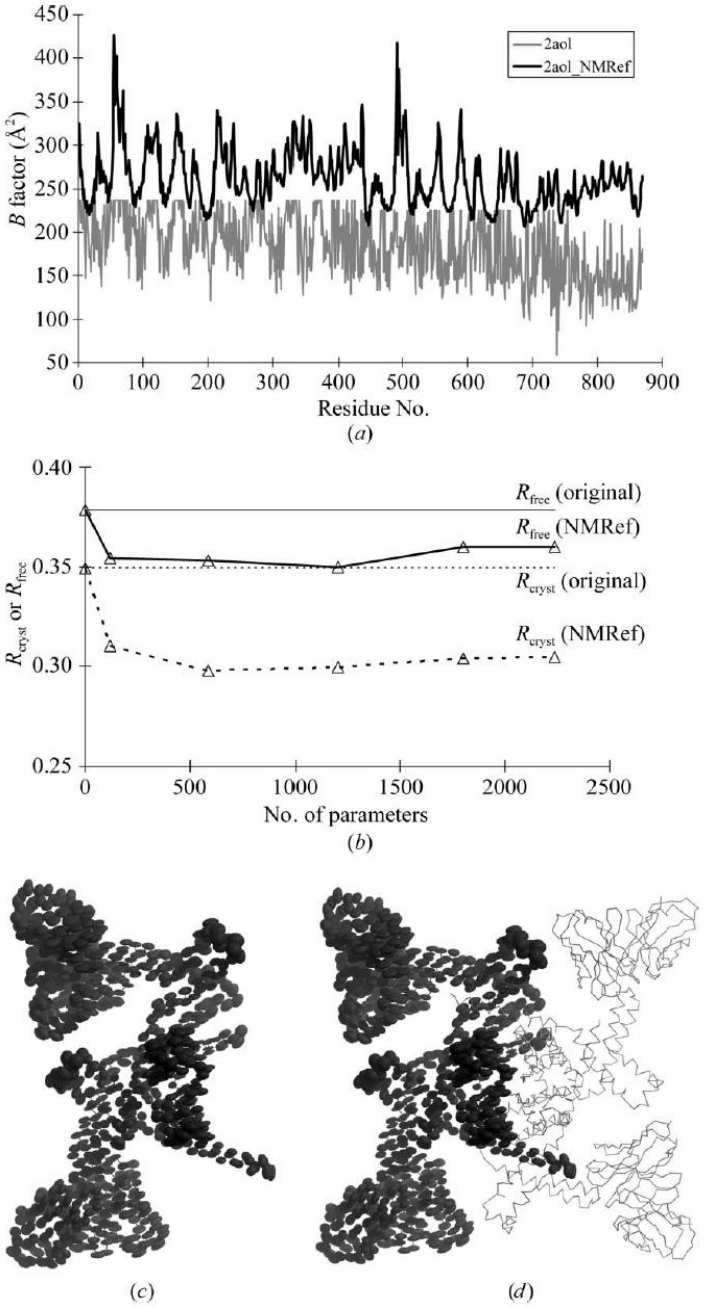

Figure 3.1 Normal mode refinement of the structure of KvAP-33H1 and Fv complex (PDB accession code: 2A0L). a) *B*-factor profile for all C_α atoms in the reported structure (light line) and normal-mode refined structure (dark line). b) The R_{cryst}- (dashed line) and R_{free}- (solid line) factors with respect to the number of parameters used in the normal-mode refinement. Original values were plotted as a dashed line and a solid line for R_{cryst}- and R_{free}-factors, respectively. c) Ellipsoids for the C_α atoms captured from the anisotropic thermal factors derived from normal-mode refinement for the asymmetric unit. d) The biological unit was shown together with the ellipsoids for the asymmetric unit. Made with 30% probability.

3.1.2. Structure of the CorA Mg^{2+} transporter at 3.9 Å resolution

The reported structure of the CorA Mg^{2+} transporter (PDB accession code: 2BBJ) was determined against the diffraction data in the range of 20.0~3.90 Å with 27,320 unique reflections, and the model contained 13,805 non-hydrogen protein atoms (*56*). The published structural model had R_{cryst}- and R_{free}-factors of 36.3% and 40.6%, respectively. The average *B*-factor for the structural model was 163 $Å^2$, with an almost flattened *B*-factor distribution (Figure 3.2a, light line). Apparently, the flexibility of the protein was not appropriately reflected by the modeled *B*-factors due to the very low data-to-parameter ratio.

Prior to the normal-mode refinement, the *R*-factors for the original PDB structure was recalculated using REFMAC5 (*97*), yielding R_{cryst}- and R_{free}-factors of 34.8% and 39.5%, respectively. Since the asymmetric unit of the deposited structure of CorA Mg^{2+} transporter was in a pentameric organization, which is also its biological form (*56*), the normal modes were directly calculated on the pentamer, with the optimal combination of the cutoff and stiffness values of 13 Å and 3. The refinement yielded the lowest R_{cryst}- and R_{free}-factors of 29.6% and 35.8% (Figure 3.2b; Table 3.1). In relation to the original structure, the R_{cryst}- and R_{free}-factors were decreased by 5.2% and 3.7%, respectively. This improvement was achieved using 62 modes, equivalent to 1980 thermal parameters that are about 1/7 of the thermal parameters used in the original structural refinement.

The residues in the original model had a distribution of 90.0 and 2.30% in the Ramachandran most favored and outlier regions, respectively (*56*). For the

normal-mode refined model, the distribution became 87.1% and 2.79% in the corresponding regions, which is comparable to that of the original model (Table 3.2).

The *B*-factor distribution of the normal-mode refined structural model is shown in Figure 3.2a (black line). Mapping it to the structure revealed much smaller *B*-factors for the transmembrane helices and higher *B*-factors for the peripheral region of the cytoplasmic domain (Figure 3.2c and Figure 3.2d), consistent with the overall architecture of the transporter.

Moreover, the multi-group TLS refinement was conducted on the normal-mode refined structure, and further reduced the R_{cryst}- and R_{free}-factors by 0.3% and 1.2%, respectively, using 15 groups per chain (1500 parameters) (Table 3.1).

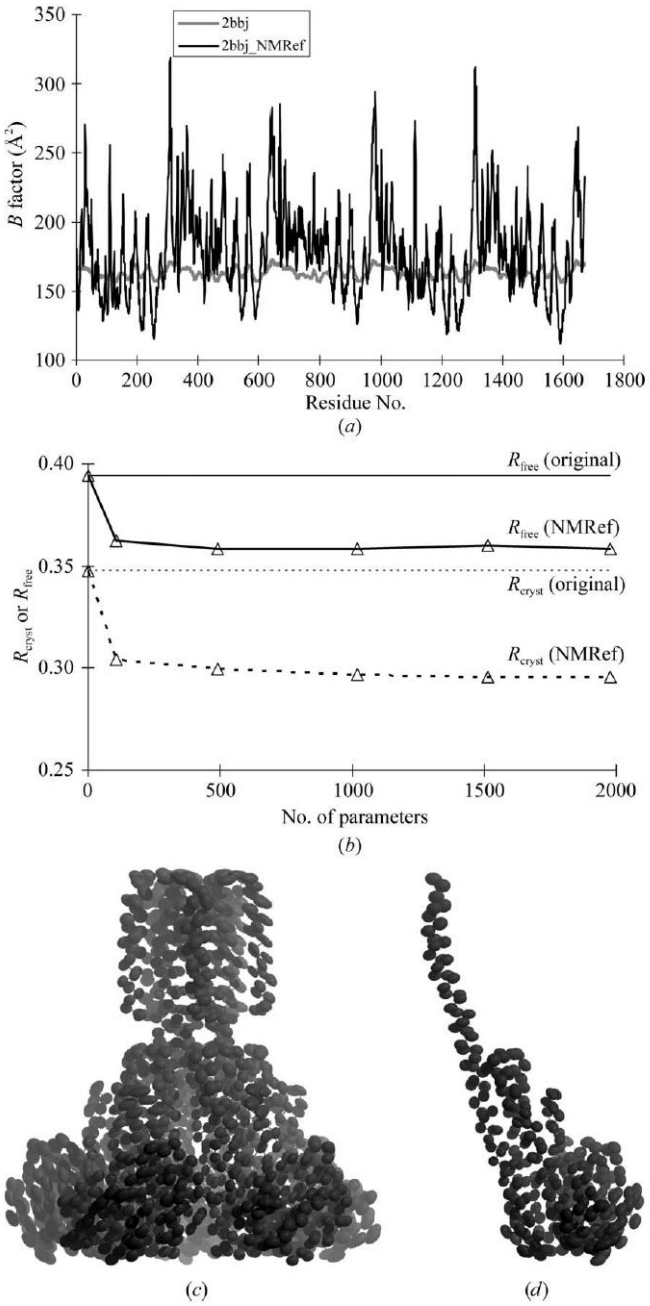

Figure 3.2 Normal mode refinement of the structure of CorA Mg^{2+} transporter (PDB accession code: 2BBJ). a) *B*-factor profile for all C_α atoms in the original structure (light line) and normal-mode refined structure (black line). b) The R_{cryst}- (dashed line) and R_{free}- (solid line) factors with respect to the number of parameters used in the normal-mode refinement. Original values were plotted as a dashed line and a solid line for R_{cryst}- and R_{free}- factors, respectively. c) Ellipsoids for the C_α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. d) Ellipsoids for a single chain of the homo-pentamer were shown. Made with 30% probability.

3.1.3. Structure of yeast Sec13/31 vertex element of the COPII vesicular coat at 3.3 Å resolution

The structure of yeast Sec 13/31 vertex element (PDB accession code: 2PM9) had 5,170 non-hydrogen protein atoms and 28 water molecule (*57*). This structure was refined against diffraction data in the range of 40.0~3.30 Å (18,996 unique reflections) to obtain R_{cryst}- and R_{free}-factors of 25.1% and 30.5%, respectively. Although the average *B*-factor of the model is 79 Å^2, some regions have *B*-factors of over 170 Å^2, indicating a rather high local flexibility (Figure 3.3a, light line).

The original model was input to REFMAC5 (*97*), yielding R_{cryst}- and R_{free}-factors of 23.9% and 28.7%, respectively. The normal-mode calculation was carried out on the asymmetric unit that was also its biological form, and the best combination of cutoff and stiffness values were 13 Å and 100, respectively. The lowest R_{cryst}- and R_{free}-factors of 22.4% and 25.9% were obtained (Figure 3.3b), representing decreases of 1.5% and 2.6% in R_{cryst}- and R_{free}-factors, respectively (Table 3.1). The improvement was achieved using five normal modes (42 parameters), corresponding to a 100-fold reduction in thermal parameters compared with the original structural model.

The residues in the original model have a distribution of 74.7% and 8.38% in the Ramachandran most favored and outlier regions, respectively. The distribution became 83.7% and 6.25% for the residues in the normal-mode refined structure, indicating a substantial improvement in geometry over the original model.

The isotropic *B*-factors for each C_α-atom, converted from the anisotropic *B*-factors of the normal-mode model, are shown in Figure 3.3a (dark line). Mapping the anisotropic *B*-factors on the structure (Figure 3.3c) revealed highly ordered core residues and much more mobile regions on the surface of the protein complex.

The subsequent multi-group TLS refinement on the normal-mode refined structure resulted in further improvement. The lowest R_{cryst}- and R_{free}-factors were 21.1% and 24.3%, respectively, when each chain was divided into 15 groups (600 parameters). These values correspond to decreases of 1.3% and 1.6% in R_{cryst}- and R_{free}-factors, respectively. In addition, the geometry was also improved slightly with 83.8% and 5.95% of residues in the Ramachandran favored and outlier regions, respectively (Table 3.2).

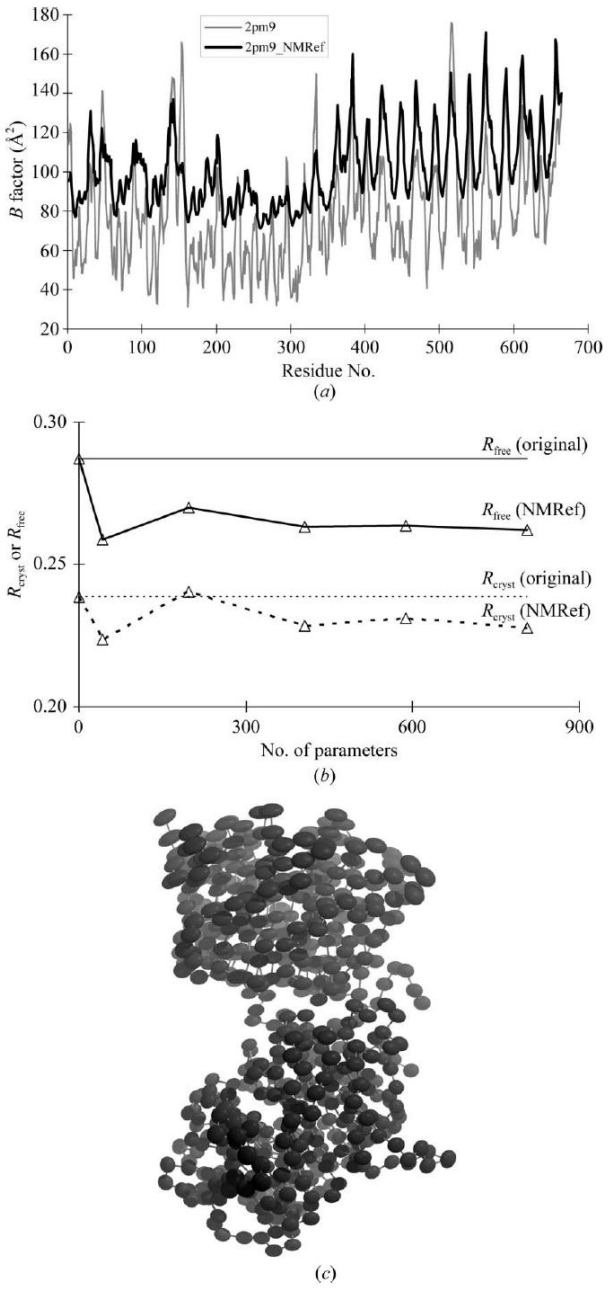

Figure 3.3 Normal mode refinement of structure of yeast Sec13/31 vertex element of the COPII vesicular coat (PDB accession code: 2PM9). a) *B*-factor profile for all C_α atoms in the original structure (light line) and normal-mode refined structure (dark line). b) The R_{cryst}- (dashed line) and R_{free}- (solid line) factors with respect to the number of parameters used in the normal-mode refinement. Original values were plotted as a dashed line and a solid line for R_{cryst}- and R_{free}-factors, respectively. c) Ellipsoids for the C_α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. Made with 30% probability.

3.1.4. Structure of HIV-1 Reverse transcriptase (RT) in complex with Jassen-R185545 at 3.0 Å resolution

The original structure of HIV-1 RT complex (PDB accession code: 1SUQ) contained 8,055 non-hydrogen protein atoms and 29 heterogen atoms, and was refined against diffraction data in the range of 19.99~3.0 Å with 28,520 unique reflections (*58*). The final model was refined using CNS (*113*) with R_{cryst}- and R_{free}-factors of 26.2% and 32.5%, respectively. This structure has an average *B*-factor of 84 Å^2. However, the large *B*-factors in some regions (over 100 Å^2) indicated a high-degree local structural flexibility (Figure 3.4a, light line).

The original model was input to REFMAC5 to recalculate the *R*-factors, yielding R_{cryst}- and R_{free}-factors of 26.0% and 31.6%, respectively. The normal-mode calculation was directly performed on the asymmetric unit, and the optimal combination of cutoff and stiffness values were 13 Å and 30. With the inclusion of 33 modes (588 parameters), the refinement yielded the lowest R_{cryst}- and R_{free}-factors of 24.0% and 29.0%, respectively (Figure 3.4b), representing a decrease of 2.0% in R_{cryst}-factor and 2.6% in R_{free}-factors (Table 3.1).

The residues of the original structural model had a distribution of 78.9% and 4.31% in the Ramachandran most favored and outlier regions, respectively. The distribution became 85.5% and 2.77% in corresponding regions for the normal-mode refined structure, indicating a substantial improvement of geometry over the original model (Table 3.2).

Overall, the isotropic *B*-factors converted from the anisotropic *B*-factors of the normal-mode model agreed well with those in the original structure (Figure 3.4a, dark line). However, a number of regions near the surface of the structure (Figure 3.4c) appeared to be of much higher mobility, consistent with the overall architecture of the enzyme.

In addition, the subsequent multi-group TLS refinement did not further improve the refinement (Table 3.1).

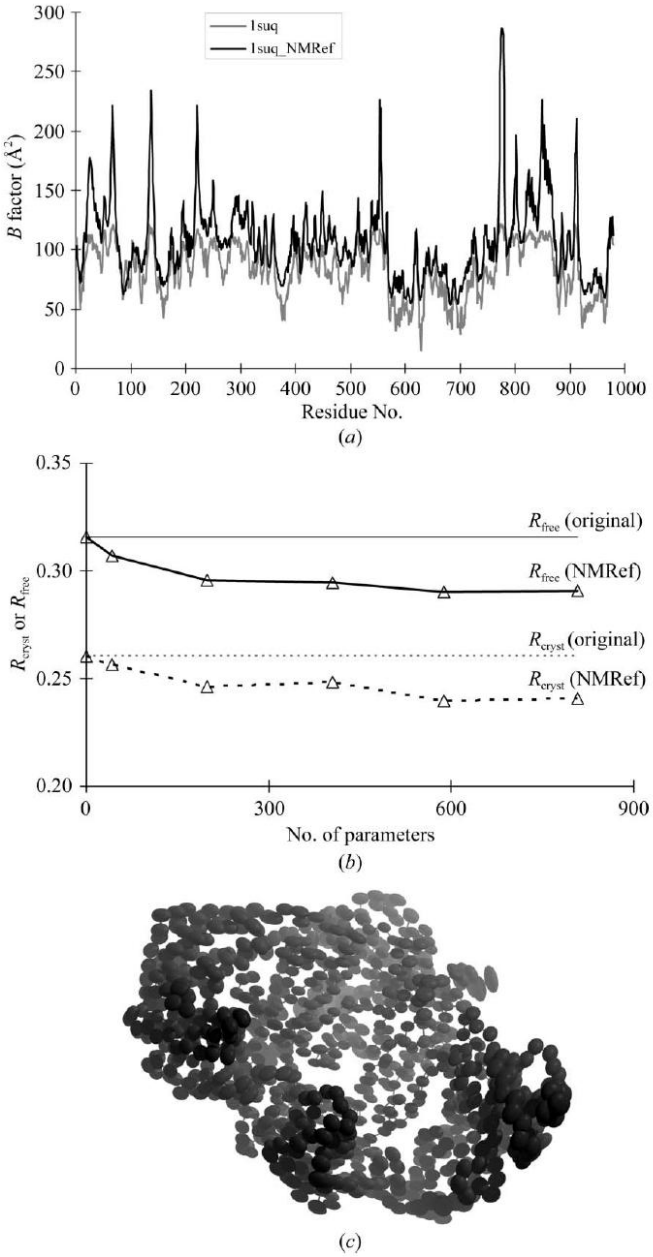

Figure 3.4 Normal mode refinement of the structure of HIV-1 reverse transcriptase (RT) in complex with Jassen-R185545 (PDB accession code: 1SUQ). a) *B*-factor profile for all C_α atoms in the original structure of HIV-1 Reverse transcriptase (RT) in complex with Jassen-R185545 (PDB accession code: 1SUQ) (light line) and normal-mode refined structure (dark line). b) The R_{cryst}- (dashed line) and R_{free}- (solid line) factors with respect to the number of parameters used in the normal-mode refinement. Original values were plotted as a dashed line and a solid line for R_{cryst}- and R_{free}-factors, respectively. c) Ellipsoids for the C_α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. Made with 50% probability.

3.1.5. The open structure of mechanosensitive channel at 3.45 Å resolution

The crystal structure of mechanosensitive channel MscS in its open form (PDB accession code: 2VV5) was determined to 3.45 Å against 46,433 unique reflections in the range of 37~3.45 Å (*59*). The reported model had 13,692 non-hydrogen protein atoms, and was refined in REFMAC5 with R_{cryst}- and R_{free}-factors of 29.3% and 31.2%, respectively. In the original structure determination, four TLS groups, one for TM1 and TM2, one for TM3a and TM3b, one for the first cytoplasmic domain and one for the second and third cytoplasmic domains, were used to facilitate the refinement. This system provided a case to test whether the normal-mode-based refinement is able to further improve TLS-refined structures.

The original model was first input to REFMAC5 to calculate *R*-factors, yielding R_{cryst}- and R_{free}-factors of 29.3% and 31.2%, respectively, the same as the published values. The normal-mode calculation was directly performed on the heptamer in the asymmetric unit, and the optimal combination of cutoff and stiffness values were 13 Å and 3. With the inclusion of 15 modes (147 parameters), the normal-mode refinement yielded the lowest R_{cryst}- and R_{free}-factors of 26.1% and 26.8% (Figure 3.5b), representing decreases of 3.2% and 4.4% in R_{cryst}- and R_{free}-factor, respectively (Table 3.1), with nearly 100-fold reduction in the number of thermal parameters.

The isotropic *B*-factor profile converted from the anisotropic one in the normal-mode refined model is shown in Figure 3.5a (dark line). Overall, the

normal-mode model appeared to be more reasonable for the resolution of this structure (Figure 3.5c and Figure 3.5d).

The residues in the original structure have a distribution of 92.4% and 2.75% in the Ramachandran most-favored and outlier regions, respectively. For the normal-mode refined mode, the distribution became 93.8% and 2.35% in the corresponding regions, which is comparable to that of the original model (Table 3.2). However, the number of bad contacts decreased to 7 in the normal-mode refined model from 37 in the original model, indicating a slight improvement of geometry.

The subsequent multi-group TLS refinement was carried out based on the normal-mode refined model. When each chain was divided into 3 groups (totally 420 parameters), the R_{cryst}-factor was further reduced by 0.3%, while the R_{free}-factor remained almost the same (Table 3.1). The distribution in the Ramachandran plot of the multi-group TLS-refined model did not differ much from the normal-mode refined model (Table 3.2).

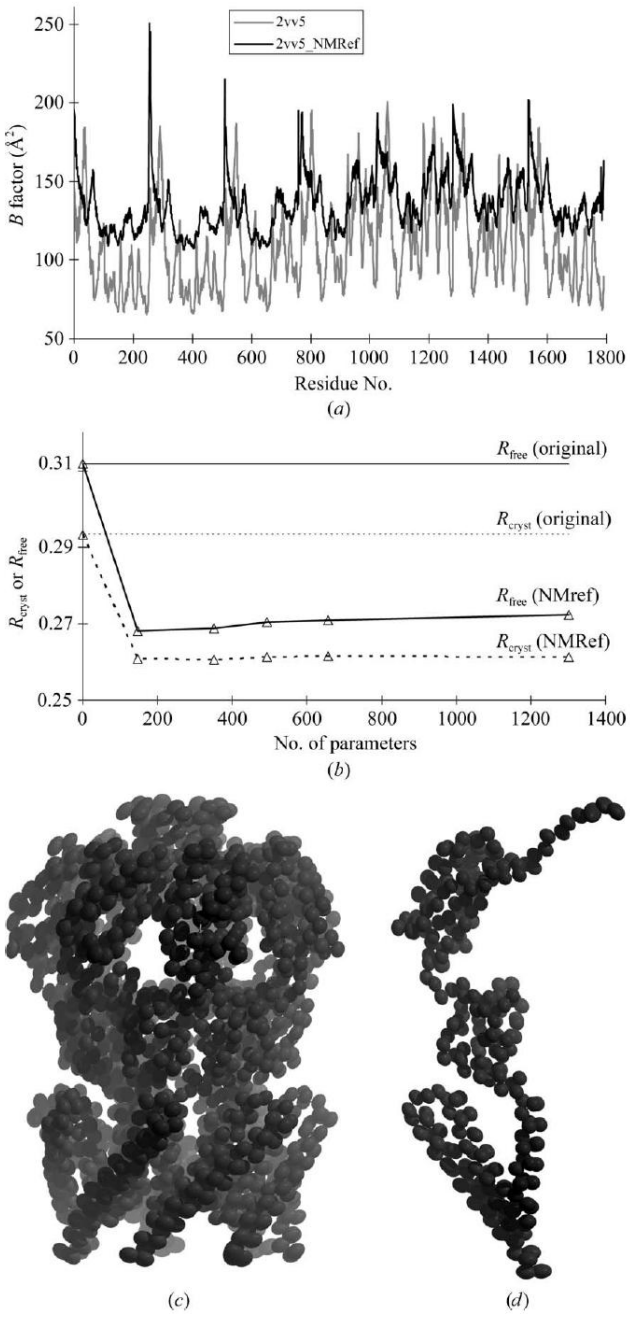

Figure 3.5 Normal mode refinement of the structure of *E.Coli* mechanosensitive channel MscS (PDB accession code: 2VV5). a) *B*-factor profile for all C_α atoms in the original structure (light line) and normal-mode refined structure (black line). b) The R_{cryst}- (dashed line) and R_{free}- (solid line) factors with respect to the number of parameters used in the normal-mode refinement. Original values were plotted as a dashed line and a solid line for R_{cryst}- and R_{free}-factors, respectively. c) Ellipsoids for the C_α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. d) Ellipsoids on a single chain of the homo-heptamer. Made with 50% probability.

3.1.6. Structure of influenza virus neuraminidase complexed with its antibody at 3.0 Å resolution

The structure of influenza virus neuraminidase complexed with its antibody was determined to 3.0 Å resolution from diffraction data collected at room temperature (22 ^{o}C) (PDB accession code: 2AEQ) (*60*). The diffraction data had a total of 23,223 unique reflections, just about four times the total of non-hydrogen atoms in the final structural model (4760 protein atoms and 142 heterogen atoms). Probably to improve the data-to-parameter ratio, the model was refined using CNS (*113*) with a constant isotropic *B*-factor of 32.64 Å^2 for all atoms and had R_{cryst}- and R_{free}-factors of 26.7% and 31.2%, respectively (Figure 3.6a, light line).

The structural model was first input to REFMAC5, yielding R_{cryst}- and R_{free}-factors of 28.5% and 29.1%, respectively (Table 3.1). Although neuraminidase is naturally a homotetramer, there is only one monomer in the asymmetric unit of this structure. Thus, normal-mode calculations were performed on the tetramer with the optimal combination of cutoff distance of 13 Å and stiffness of 3. The plot of R_{cryst}- and R_{free}-factors as a function of the number of parameters used in the refinement suggested that inclusion of 33 modes (totally 588 parameters) resulted in the lowest R_{cryst}- and R_{free} -factors (24.5% and 27.4%, respectively) (Figure 3.6b). This represents decreases of 4.0% and 1.7% in R_{cryst}- and R_{free}-factors, respectively, achieved by using ~10-fold fewer thermal parameters than in the original structural refinement (Table 3.1).

In the Ramachandran plot, the original model had 85.9% and 3.64% residues in the Ramachandran most favored and outlier regions, respectively. The normal-mode-refined structural model had a slightly improved geometry, with 88.4% and 2.64% of the residues in these regions, respectively (Table 3.2).

A second structure of neuraminidase was also reported, which was determined from a frozen crystal but otherwise the same sample at 2.1 Å resolution (PDB accession code: 2AEP) (*60*). As a way to compare, Fig.6a shows the isotropic-equivalent *B*-factors for C_α-atoms in the normal-mode refined structure, the constant *B*-factors in the original structure (PDB accession code: 2AEQ) at 3.0 Å resolution, and the isotropic *B*-factors in the 2.1 Å structure (PDB accession code: 2AEP). It is clear that the profile of *B*-factor in the normal-mode refined structure matches with that of the 2.1 Å structure very well; however, higher absolute values, consistent with its lower resolution, were observed. Thus, it is evident that the normal-mode refinement delivers a reasonable *B*-factor profile for 2AEQ that is otherwise not practical in conventional refinement. In addition, the ellipsoids mapped to the structure revealed higher structural flexibility for the outer regions of the tetrameric structure, as expected for protein structures (Figure 3.6c and Figure 3.6d).

With the *B*-factor profile in the normal-mode refined structure, automated multi-group TLS refinement was conducted to see if further decreases in *R*-factors can be achieved. Indeed, the R_{cryst}- and R_{free}-factors were further reduced by 1.5% and 0.4% using 15 groups per chain (equals to 900 parameters) (Table 3.1).

Moreover, the normal-mode and TLS model had a distribution as 89.9% and 1.82% in the Ramachandran favored and outlier regions, respectively, indicating a slight further improvement of the geometry.

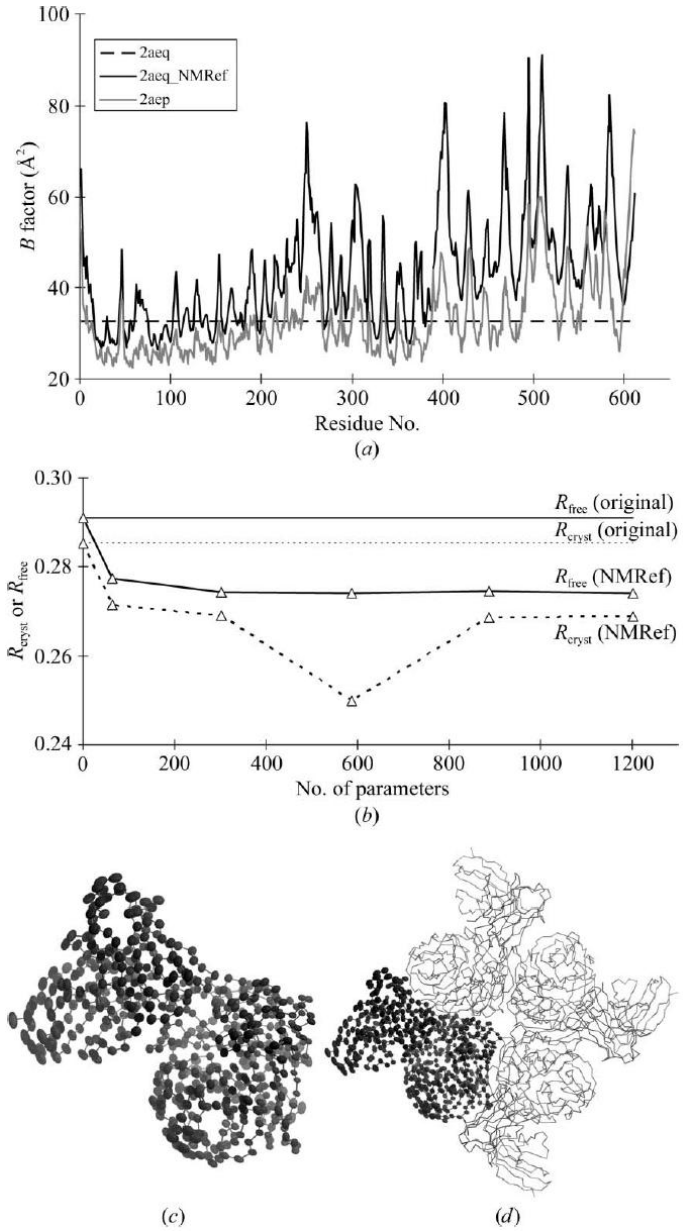

Figure 3.6 Normal mode refinement of the structure of influenza virus neuraminidase and its antibody (PDB accession code: 2AEQ). a) *B*-factor profile for all C_α atoms in the original structure (dashed line), its high-resolution structure (PDB acession code: 2AEP) (light solid line), and normal-mode refined structure (dark line). b) The R_{cryst}- (dashed line) and R_{free}- (solid line) factors with respect to the number of parameters used in the normal-mode refinement. Original values were plotted as a dashed line and a solid line for R_{cryst}- and R_{free}-factors, respectively. c) Ellipsoids for the C_α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. d) The biological unit was shown together with the ellipsoids for the asymmetric unit. Made with 50% probability.

3.1.7. Structure of chicken embryo lethal orphan adenovirus major coat protein, hexon at 3.9 Å resolution

The reported model of hexon (PDB accession code: 2INY) had R_{cryst}- and R_{free}-factors refined in CNS of 37.2% and 41.6%, respectively (*61*). The diffraction data in the range of 47.06 ~ 3.90 Å had 14,124 unique reflections, whereas the final structural model had 7,523 non-hydrogen protein atoms. Thus, the data-to-parameter ratio was rather low. Possibly for this reason, the structure was refined by constraining all *B*-factors to 30.0 Å^2 (Figure 3.7a, light line). According to the authors, various efforts were made to further refine the structure, but all proved futile.

Prior to normal-mode based refinement, the *R*-factors for the original model were recalculated using REFMAC5, yielding R_{cryst}- and R_{free}-factors of 34.4% and 43.3%, respectively. The normal-mode calculation was performed on the trimeric biological unit, and the optimal combination of the cutoff distance and stiffness values were 20 Å and 3. Only the portion of eigenvectors corresponding to the monomeric asymmetric unit was included in normal-mode refinement (Figure 3.7b). With the inclusion of 12 modes (105 parameters), the refinement yielded the lowest R_{cryst}- and R_{free}-factors of 31.4% and 40.1%, representing decreases of 3.0% and 3.2% in R_{cryst}- and R_{free}-factors, respectively. The improvement was achieved by using 70-fold fewer thermal parameters than in the original structural refinement.

The residues in the original model had a distribution of 59.2% and 14.5% in the Ramachandran most favored and outlier regions, respectively. After normal-

mode refinement, the distribution became 65.5% and 8.87% in corresponding regions, indicating a substantial improvement in geometry over the original model (Table 3.2).

The isotropic *B*-factors converted from the anisotropic *B*-factors in the normal-mode refined model were compared with those of the original model in Figure 3.7a (dark line). Although the majority of the structure has a rather low *B*-factor, there are a number of regions of extremely high mobility (Figure 3.7a). As expected, these regions are located far away from other structural components in native trimeric structure (Figure 3.7c and Figure 3.7d).

However, the subsequent multi-group TLS refinement on the normal-model refined structure did not result in further decreases in *R*-factors (Table 3.1). Thus, in this case, no gain was obtained by further applying TLS refinement.

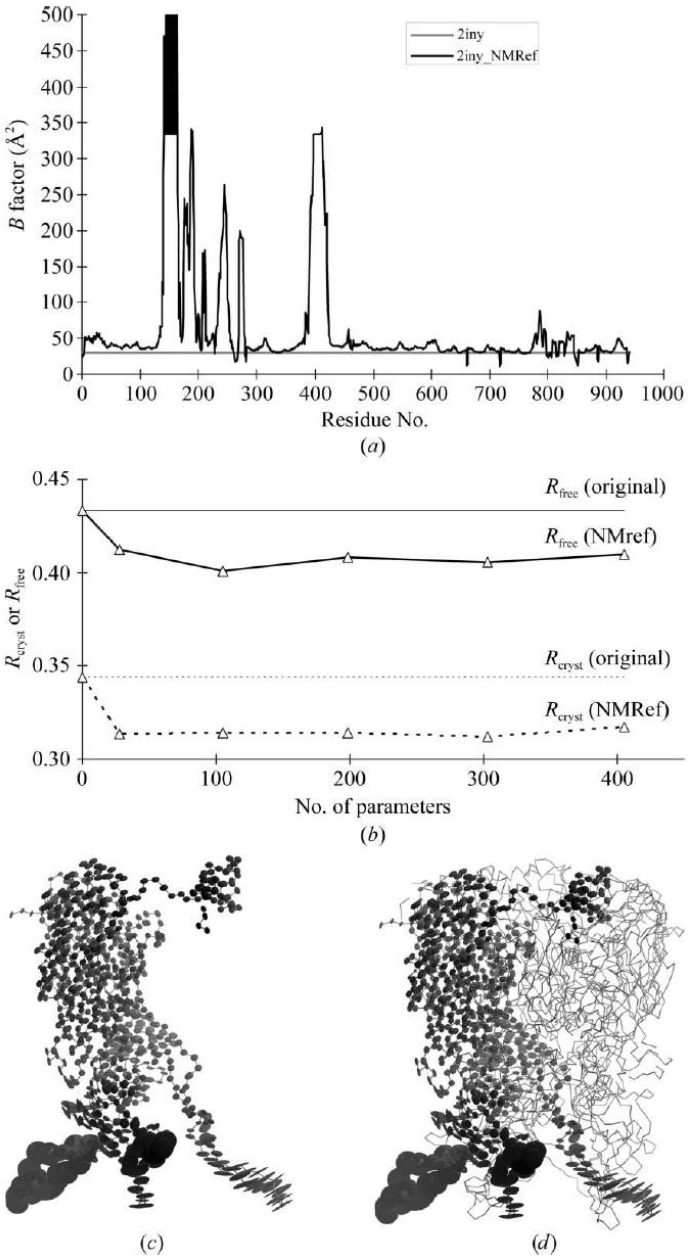

Figure 3.7 Normal mode refinement of the structure of the chicken embryo lethal orphan adenovirus major coat protein (PDB accession code: 2INY). a) *B*-factor profile for all C_α atoms in the original structure (light line) and normal-mode refined structure (dark line). b) The R_{cryst}- (dashed line) and R_{free}- (solid line) factors with respect to the number of parameters used in the normal-mode refinement. Original values were plotted as a dashed line and a solid line for R_{cryst}- and R_{free}-factors, respectively. c) Ellipsoids for the C_α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. d) The biological unit was shown together with the ellipsoids for the asymmetric unit. Made with 50% probability.

3.1.8. Structure of *Deinococcus radiodurans* RecR and RecO (drRecOR) at 3.8 Å resolution

The original structural model of drRecOR (PDB accession code: 2V1C) had a total of 4,778 non-hydrogen protein atoms and three zinc atoms, and was refined against the diffraction data in the range of 3.80~71.25 Å (with 6,850 unique reflections) (*62*). Due to the very low data-to-parameter-ratio, the reported refinement was terminated after a single round of rigid-body refinement and manual rebuilding with R_{cryst}- and R_{free}-factors of 45.8% and 44.3%, respectively. The final model has a constant *B*-factor of 96 Å^2 (Figure 3.8a, light line).

Prior to the normal-mode based refinement, the original model was input to REFMAC5, yielding R_{cryst}- and R_{free}-factors of 43.1% and 45.7%, respectively. The normal-mode calculation was performed on the dimeric biologic unit and the best combination of cutoff and stiffness values were 13 Å and 100. The portion of eigenvectors corresponding to the monomeric asymmetric unit was included in the normal-mode refinement using a different number of normal modes (Figure 3.8b). The lowest R_{cryst}- and R_{free}-factors were 28.2% and 38.0%, respectively, when the eight lowest normal modes (equivalent to 63 thermal parameters) were used. Compared with the original model, this result represents decreases of 14.9% and 7.7% in R_{cryst}- and R_{free}-factors, respectively.

The residues in the original model are distributed with 90.8% and 2.26% in the Ramachandran most favored and outlier regions, respectively. The distribution became 77.1% and 5.33%, respectively, for the normal-mode-refined structure.

This result may be ascribed to the fact, at least in part, that the original model for drRecOR was constructed based on the structures of drRecO and drRecR determined at much higher resolution without further positional refinement. However, the number of bad contacts decreased to 23 in the normal-mode refined structure from 58 in the original model.

The *B*-factors from the normal-mode refined model (Figure 3.8a) suggest realistic flexibility variations along the structure, with more ordered region inside and more flexible regions outside (Figure 3.8c and Figure 3.8d).

In addition, the subsequent multi-group TLS refinement on the normal-mode refined structure did not make any remarkable improvement. Although the R_{free}-factor was decreased to 37.9% with 10 TLS groups per chain, the R_{cryst}-factor was increased to 29.1% (Table 3.1).

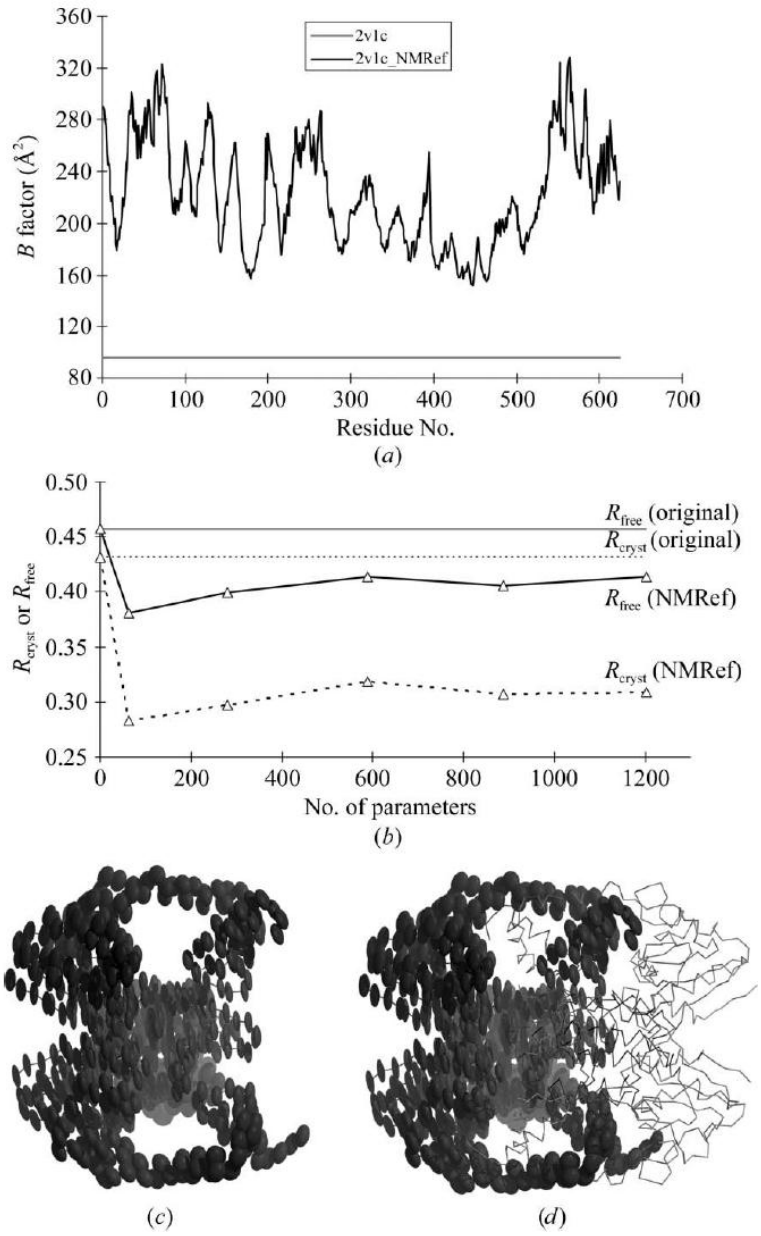

Figure 3.8 Normal mode refinement of the structure *Deinococcus radiodurans* RecR and RecO (PDB accession code: 2V1C). a) *B*-factor profile for all C_α atoms in the original structure of (light line) and normal-mode refined structure (dark line). b) The R_{cryst}- (dashed line) and R_{free}- (solid line) factors with respect to the number of parameters used in the normal-mode refinement. Original values were plotted as a dashed line and a solid line for R_{cryst}- and R_{free}-factors, respectively. c) Ellipsoids for the C_α atoms captured from the anisotropic thermal factor derived from normal-mode refinement for the asymmetric unit. d) The biological unit was shown together with the ellipsoids for the asymmetric unit. Made with 50% probability.

3.2. Role of Phe95→Tyr mutation in binding affinity of BHA

3.2.1. Expression of recombinant wild-type and mutant BHA glycoproteins

The recombinant wild-type BHA and its mutants (single mutants: Phe95→Tyr, Asn194→Asp and double mutant Phe95→Tyr/Asn194→Asp) were expressed in mammalian cells CV-1 using vaccinia viral system (kindly provided by Dr. Bernard Moss) (*98*) and purified to high purity (see Materials and Methods). For the Asn194→Asp and Phe95→Tyr/Asn194→Asp mutants, the glycosylation site was depleted in order to mimic the situation in which influenza BHA lost its glycosylation site at this position in some field isolates (*79*, *80*) or due to egg adaptation (*72*–*78*). Consequently, the molecular weights of the HA0 monomer for these two mutants were about three kilo-daltons smaller than that of wild-type BHA as shown on the SDS-PAGE (Figure 3.9).

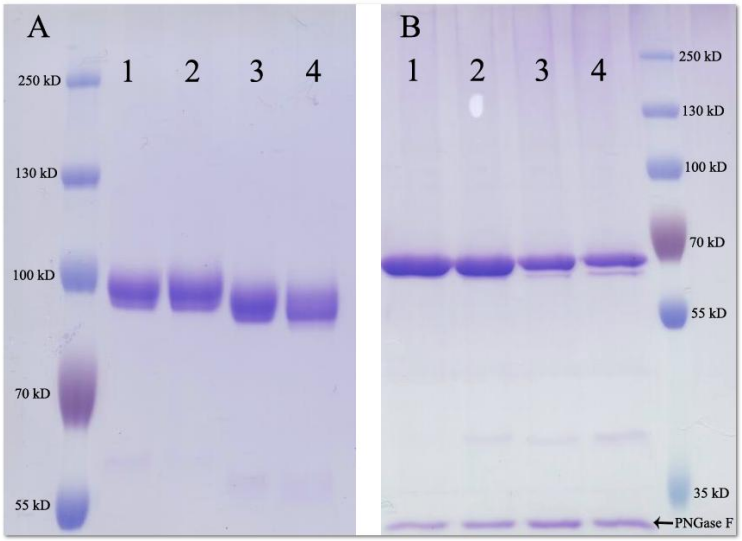

Figure 3.9 SDS-PAGE analysis of recombinant wild type BHA and its mutants. Shown are wild type BHA (1), mutant Phe95→Tyr (2), Asn194→Asp (3) and Phe95→Tyr/ Asn194→Asp (4). (A) Recombinant proteins after gel filtration purification. (B) Purified proteins after treatment with PNGase F to remove the glycans.

3.2.2. The mutation Phe95→Tyr universally enhanced the binding affinity of BHA

To investigate the impact of Phe95→Tyr mutation on the receptor binding capability of BHA, a glycan microarray analysis on wild-type BHA and each of its mutants (Figure 3.10) is carried out. Out of the more than 400 glycans on the microarray chips, a total of 97 glycans showed significant binding in at least one of the four protein samples (wild-type BHA and its mutants) (Table 3.3). The fluorescence intensity is shown for each of these glycans for each of the tested proteins in Figure 3.10A~D in which the glycans are grouped according to the presence or absence of a terminal sialic acid group and the linkage between the

sialic acid group. Clearly for wild-type BHA, it has an overall weak binding for the majority of glycans, although it does exhibit a preference towards α(2,6)-linked human-like receptor analogs (red in Figure 3.10), consistent with the abundance of α(2,6)-linked sialic acids in human upper respiratory tract where the wild-type influenza B virus infects its human host. Along the same line, the mutant Asn194→Asp, which causes the loss of a glycosylation site at HA_1 194∼196, as frequently found in egg-adapted variants (*72*–*78*), resulted in a preferential binding to α(2, 3)-linked avian-like glycan, and the overall fluorescence intensity is substantially increased compared to the wild-type BHA.

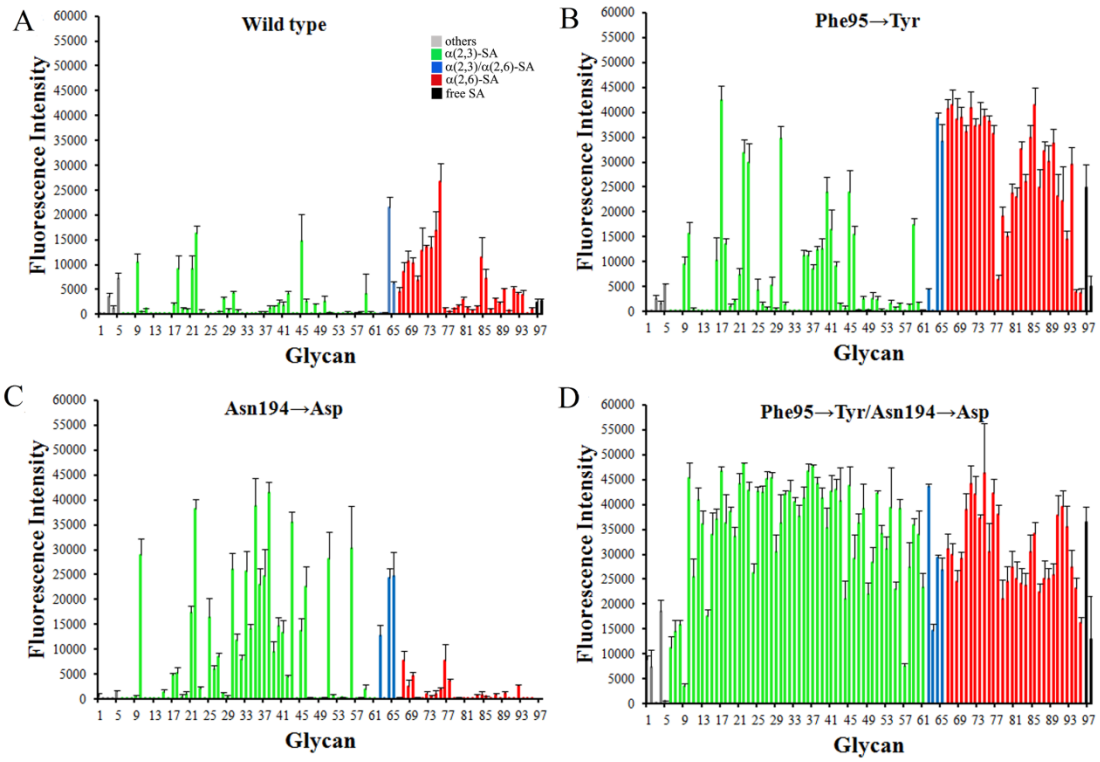

Figure 3.10 The mutation Phe95→Tyr promotes the binding of BHA to synthetic glycans in glycan array assays. The results of glycan array assays are shown for (A) wild type, (B) mutant Phe95→Tyr; (C) mutant Asn194→Asp; and (D) mutant Phe95→Tyr/Asn194→Asp. Please refer to Table 3.3 for the full name of each glycan. The glycans are categorized into α2-3-linked sialosides (■), glycans containing both α2-3 and α2-6 sialosides (■) α2-6-linked sialosides (■), free sialic acid (■) and others (■).

Table 3.3 Full names for the glycans of which the relative fluorescence unit (RFU) are above 5000

glycan array	Glycan Structure
1	Fucα1-2[6OSO3]Galβ1-4[6OSO3]Glc-Sp0
2	Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ-Sp10
3	Neu5Acα2-8Neu5Acα2-8Neu5Acα-Sp8
4	Galβ1-4GlcNAcβ1-2Manα-Sp0
5	GlcNAcα1-4Galβ1-4GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAcβ-Sp0
6	Fucα1-2Galβ1-3GalNAcβ1-4(Neu5Acα2-3)Galβ1-4Glcβ-Sp0
7	Fucα1-2Galβ1-3GalNAcβ1-4(Neu5Acα2-3)Galβ1-4Glcβ-Sp9
8	Galβ1-3GalNAcβ1-4(Neu5Acα2-3)Galβ1-4Glcβ-Sp0
9	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAc-Sp0
10	Neu5Acα2-3Galβ1-3GalNAcα-Sp8
11	Neu5Acα2-3Galβ1-3GalNAcα-Sp14
12	Neu5Acα2-3(GalNAcβ1-4)Galβ1-4GlcNAcβ-Sp0
13	Neu5Acα2-3(GalNAcβ1-4)Galβ1-4GlcNAcβ-Sp8
14	Neu5Acα2-3(GalNAcβ1-4)Galβ1-4Glcβ-Sp0
15	Neu5Acα2-3(Neu5Acα2-3Galβ1-3GalNAcβ1-4)Galβ1-4Glcβ-Sp0
16	Neu5Acα2-3GalNAcα-Sp8
17	Neu5Acα2-3GalNAcβ1-4GlcNAcβ-Sp0
18	Neu5Acα2-3Galβ1-3[6OSO3]GlcNAc-Sp8
19	Neu5Acα2-3Galβ1-3(Fucα1-4)GlcNAcβ-Sp8
20	Neu5Acα2-3Galβ1-3(Fucα1-4)GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAcβ-Sp0
21	Neu5Acα2-3Galβ1-3(Neu5Acα2-3Galβ1-4)GlcNAcβ-Sp8
22	Neu5Acα2-3Galβ1-3[6OSO3]GalNAcα-Sp8
23	Neu5Acα2-3Galβ-Sp8
24	Neu5Acα2-3Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glcβ-Sp0
25	Neu5Acα2-3Galβ1-3GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0
26	Neu5Acα2-3Galβ1-3GlcNAcβ-Sp0
27	Neu5Acα2-3Galβ1-3GlcNAcβ-Sp8
28	Neu5Acα2-3Galβ1-4[6OSO3]GlcNAcβ-Sp8
29	Neu5Acα2-3Galβ1-4(Fucα1-3)[6OSO3]GlcNAcβ-Sp8
30	Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAcβ-Sp0
31	Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ-Sp0
32	Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ-Sp8
33	Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ-Sp8
34	Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp8
35	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0
36	Neu5Acα2-3Galβ1-4GlcNAcβ-Sp0
37	Neu5Acα2-3Galβ1-4GlcNAcβ-Sp8
38	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0
39	Neu5Acα2-3Galβ1-4Glcβ-Sp0

40	Neu5Acα2-3Galβ1-4Glcβ-Sp8
41	Galβ1-3(Neu5Acα2-3Galβ1-4GlcNAcβ1-6)GalNAcα-Sp14
42	Neu5Acα2-3Galβ1-3GlcNAcβ1-3Galβ1-3GlcNAcβ-Sp0
43	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ1-3GlcNAcβ-Sp0
44	Galβ1-3(Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ1-6)GalNAcα-Sp14
45	Neu5Acα2-3Galβ1-3(Neu5Acα2-3Galβ1-4GlcNAcβ1-6)GalNAcα-Sp14
46	Neu5Acα2-3Galβ1-4GlcNAcβ1-2Manα1-3(Neu5Acα2-3Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12
47	Neu5Ac(9Ac)a2-3Galβ1-4GlcNAcβ-Sp0
48	Neu5Ac(9Ac)a2-3Galβ1-3GlcNAcβ-Sp0
49	Neu5Acα2-3Galβ1-3(Fucα1-4)GlcNAcβ1-3Galβ1-3(Fucα1-4)GlcNAcβ-Sp0
50	Neu5Acα2-3-Galβ1-3(Galβ1-4(Fucα1-3)GlcNAcβ1-6)GalNAc-Sp14
51	Neu5Acα2-3Galβ1-4GlcNAcβ1-3GalNAc-Sp14
52	Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ1-3GalNAcα-Sp14
53	Neu5Acα2-3(GalNAcβ1-4)Galβ1-4GlcNAcβ1-3GalNAcα-Sp14
54	Neu5Acα2-3Galβ1-3GlcNAcβ1-3GalNAcα-Sp14
55	Neu5Acα2-3Galβ1-3GalNAcβ1-4(Neu5Acα2-8Neu5Acα2-3)Galβ1-4Glcβ-Sp0
56	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ-Sp8
57	Neu5Gcα2-3Galβ1-3(Fucα1-4)GlcNAcβ-Sp0
58	Neu5Gcα2-3Galβ1-3GlcNAcβ-Sp0
59	Neu5Gcα2-3Galβ1-4(Fucα1-3)GlcNAcβ-Sp0
60	Neu5Gcα2-3Galβ1-4GlcNAcβ-Sp0
61	Neu5Gcα2-3Galβ1-4Glcβ-Sp0
62	Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα-Sp8
63	Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα-Sp14
64	Neu5Acα2-3Galβ1-4GlcNAcβ1-2Manα1-3(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12
65	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3(Neu5Acα2-3Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12
66	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-N(LT)AVL
67	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp8
68	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12
69	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp13
70	Neu5Acα2-6GalNAcα-Sp8
71	Neu5Acα2-6GalNAcβ1-4GlcNAcβ-Sp0
72	Neu5Acα2-6Galβ1-4[6OSO3]GlcNAcβ-Sp8
73	Neu5Acα2-6Galβ1-4GlcNAcβ-Sp0
74	Neu5Acα2-6Galβ1-4GlcNAcβ-Sp8
75	Neu5Acα2-6Galβ1-4GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAcβ-Sp0
76	Neu5Acα2-6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0
77	Neu5Acα2-6Galβ1-4Glcβ-Sp0
78	Neu5Acα2-6Galβ1-4Glcβ-Sp8
79	Neu5Acα2-6Galβ-Sp8
80	Galβ1-4GlcNAcβ1-2Manα1-3(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12
81	GlcNAcβ1-2Manα1-3(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12
82	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3(Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12

83	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3(GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12
84	Neu5Acα2-6Galβ1-4GlcNAcβ1-3Galβ1-3GlcNAcβ-Sp0
85	Neu5Acα2-6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0
86	Manα1-3(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAc-Sp12
87	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAc-Sp12
88	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6Manβ1-4GlcNAcβ1-4GlcNAc-Sp12
89	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3Manβ1-4GlcNAcβ1-4GlcNAc-Sp12
90	Neu5Acα2-6GlcNAcβ1-4GlcNAc-Sp21
91	Neu5Acα2-6GlcNAcβ1-4GlcNAcβ1-4GlcNAc-Sp21
92	Neu5Acα2-6Galβ1-4GlcNAcβ1-3GalNAc-Sp14
93	Neu5Acα2-6Galβ1-3GlcNAcβ1-3(Galβ1-4GlcNAcβ1-6)Galβ1-4Glc-Sp21
94	Neu5Acα2-6Galβ1-4GlcNAcβ1-6(Fucα1-2Galβ1-3GlcNAcβ1-3)Galβ-4Glc-Sp21
95	Neu5Gcα2-6Galβ1-4GlcNAcβ-Sp0
96	Neu5Acα-Sp11
97	Neu5Acα-Sp8
Sp0	$CH_2CH_2NH_2$
Sp8	$CH_2CH_2CH_2NH_2$
Sp9	$CH_2CH_2CH_2CH_2CH_2NH_2$
Sp10	$NHCOCH_2NH$
Sp11	$OCH_2C_6H_4$-p-$NHCOCH_2NH$
Sp12	Asparagine
Sp13	Glycine
Sp14	Threonine
Sp15	Serine
Sp16	PNP ($OC_6H_4NH_2$)
Sp17	$OCH_2C_6H_4NH_2$
Sp18	$O(CH_2)_3NHCO(CH_2)_5NH_2$
Sp19	EN or NK
Sp20	GENR
Sp21	N(CH3)-O-(CH2)2-NH2
Sp22	NST
MDPLys	Mur-L-Ala-D-iGlnb-(CH2)4NH2

The color scheme used here is the same as that in Figure 3.10. The glycans are categorized into α(2,3)-linked sialosides (■), glycans containing both α(2,3) and α(2,6) sialosides (■), α(2,3)-linked sialosides (■), free sialic acid (■) and others (■). The linker structures are also listed at the end of the table.

Compared to the wild-type BHA, although Phe95→Tyr BHA shares the same preference for α(2,6)-linked human-like receptors (Figure 3.10B), the fluorescence intensity is significantly improved for α(2,3)-linked avian-like receptors (in green color), α(2,6)-linked human-like receptors (in red color), α(2,3)/α(2,6)-linked receptors (in blue color) and free sialic acids (in black color). Similarly, compared to Asn194→Asp, the Phe95→Tyr/Asn194→Asp mutant has significantly enhanced binding to all types of sialic-acid-containing receptors (Figure 3.10D). In addition, it is noteworthy that Phe95→Tyr/Asn194→Asp exhibited higher binding even for several glycans that do not contain sialic acid moiety (Table 3.3).

In order to quantify the impact of Phe95→Tyr mutation on the binding affinity of BHA, dose-dependent glycan binding assays using α(2,3)-linked avian-like receptor and α(2,6)-linked human-like receptor are performed (Table 3.4). Wild-type BHA preferentially binds to the α(2,6)-linked human-like receptor 6'SLN-LN. The mutation Phe95→Tyr simultaneously enhanced the binding for both avian-like and human-like receptors, 3'SLN-LN and 6'SLN-LN (Table 3.4). The calculated apparent binding constants for wild-type and Phe95→Tyr BHA were $5.0 \pm 4.2 \times 10^{-3}$ M and $1.3 \pm 0.4 \times 10^{-6}$ M for the 3'SLN-LN receptor, $1.5 \pm 0.8 \times 10^{-4}$ M and $1.1 \pm 0.7 \times 10^{-11}$ M for the 6'SLN-LN receptor, respectively (Table 3.4), representing an 10^3 and 10^7 increase in binding affinity as the result of Phe95→Tyr for 3'SLN-LN and 6'SLN-LN receptors. Not surprisingly, the mutation Asn194→Asp, which causes the loss of a glycosylation site at HA_1 194~196, as frequently found in egg-adapted variants (*72–78*), resulted in preferential binding to α(2,3)-linked avian-like receptor, with an

apparent binding constant of $3.0 \pm 1.1 \times 10^{-12}$ M for 3'SLN-LN, in contrast to the $6.0 \pm 0.9 \times 10^{-5}$ M binding constant for α(2,6)-linked 6'SLN-LN (Table 3.4). As seen for the single Phe95→Tyr mutant, the double mutant Phe95→Tyr/Asn194→Asp improved the binding for both α(2,3)-linked and α(2,6)-linked receptors, with apparent binding constants of $7.6 \pm 0.9 \times 10^{-12}$ M and $2.7 \pm 0.9 \times 10^{-12}$ M for 3'SLN-LN and 6'SLN-LN, respectively (Table 3.4). This corresponded to an increase of about 10^7 times in binding affinity towards human receptor analog by the Phe95→Tyr mutation when BHA does not have a glycosylation at HA_1 194~196. Thus, regardless of whether or not there is a glycosylation at HA_1 194~196, the mutation Phe95→Tyr enhances the binding of BHA for both avian and human-like receptors to a comparable affinity, resulting in minimal discrimination.

Table 3.4 Equilibrium dissociation constants of recombinant wild-type BHA protein and its mutants in binding to α(2,3) and α(2,6)-linked receptor analogs

BHA	3'SLN-LN*		
	n	K_d'	R^2
Wild type	0.49±0.05	$5.0 \pm 4.2 \times 10^{-3}$ M	0.97±0.01
Phe95→Tyr	0.90±0.01	$1.3 \pm 0.4 \times 10^{-6}$ M	0.994±0.003
Asn194→Asp	1.50±0.02	$3.0 \pm 1.1 \times 10^{-12}$ M	0.997±0.001
Phe95→Tyr/Asn194→Asp	1.31±0.01	$7.6 \pm 0.9 \times 10^{-12}$ M	0.994±0.001
BHA	6'SLN-LN*		
	n	K_d'	R^2
Wild type	0.61±0.03	$1.5 \pm 0.8 \times 10^{-4}$ M	0.96±0.01
Phe95→Tyr	1.27±0.03	$1.1 \pm 0.7 \times 10^{-11}$ M	0.995±0.001
Asn194→Asp	0.81±0.07	$6.0 \pm 0.9 \times 10^{-5}$ M	0.987±0.001
Phe95→Tyr/Asn194→Asp	1.63±0.02	$2.7 \pm 0.9 \times 10^{-12}$ M	0.970±0.001

* The apparent binding constant (K_d'), co-cooperativity factor (n) and R-square (R^2) by fitting the data to the linearized Hill equation in order to quantitatively determine the relative binding affinities of BHAs, and their absolute values should be compared only in this context.

3.2.3. Results of red blood cells binding assay

Red blood cells from different species are known to contain varying amounts of avian and/or human-like receptors. Thus, if the enhanced binding affinity of Phe95→Tyr for synthetic receptors holds true, it is expected that this mutation allows better binding for natural sialic acid receptors on many different red blood cell types. As shown in Figure 3.11, the big circles are Nickel-NTA agarose beads, the black dots are red blood cells. The wild type BHA protein is coated on the agarose beads through its His-tag binding to the Ni-NTA. If wild type BHA binds the receptors on RBC, the agarose bead (big circle) is surrounded by a layer of RBCs (black dots). One thing needed to be noted is that the quality of the agarose beads might also contribute to the heterogeneity of beads, where some beads are bare and some beads are coated with red blood cells.

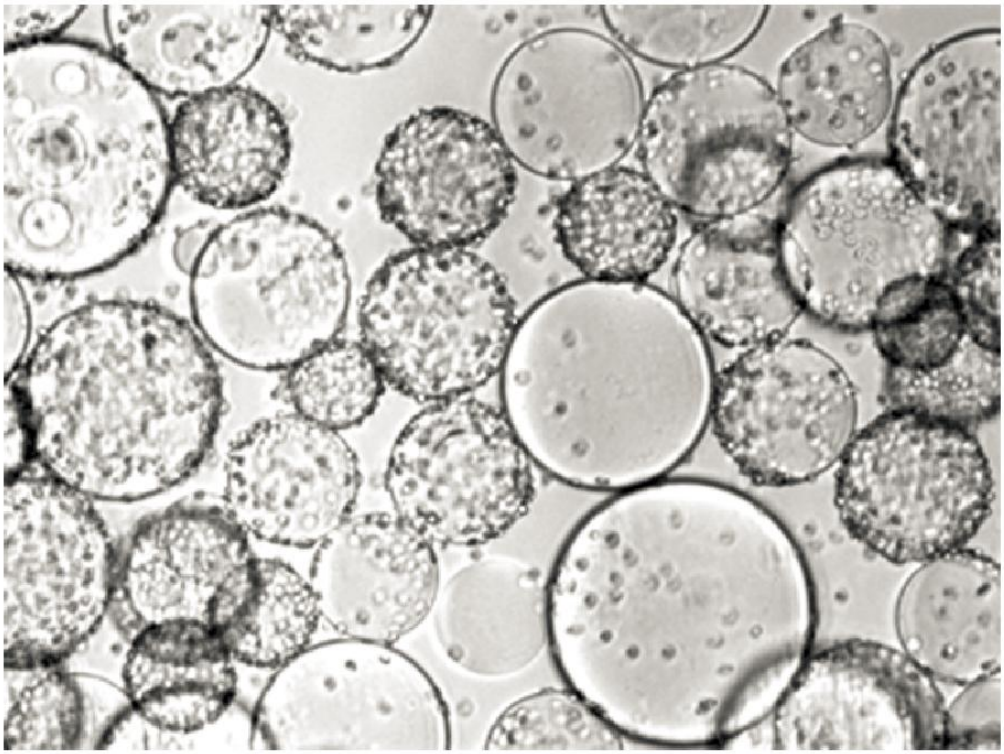

Figure 3.11 Human red blood cell (RBC) binding with wild type BHA. The big circles are Nickel-NTA agarose beads, the black dots are red blood cells. The BHA protein is coated on the agarose beads through its His-tag binding to the Ni-NTA. If BHA binds the receptors on RBC, the agarose bead (big circle) is surrounded by a layer of RBCs (black dots).

Red blood cells from eight species were tested (Figure 3.12). All four proteins were found to bind well to human red blood cells. Compared with their corresponding wild-type proteins, both Phe95→Tyr and Phe95→Tyr/Asn194→Asp mutants bound significantly better for red blood cells from chicken, turkey, and guinea pig, and somewhat better for those from rabbit and bovine. For the red blood cells from swine, the mutant Phe95→Tyr bound significantly better while Phe95→Tyr/Asn194→Asp only bound slightly better than their corresponding wild-type proteins (Figure 3.12). Therefore, the higher binding affinity of the Phe95→Tyr and Phe95→Tyr/Asn194→Asp mutants for synthetic receptors can be readily translated into a much stronger binding to natural sialic acid receptors on red blood cells.

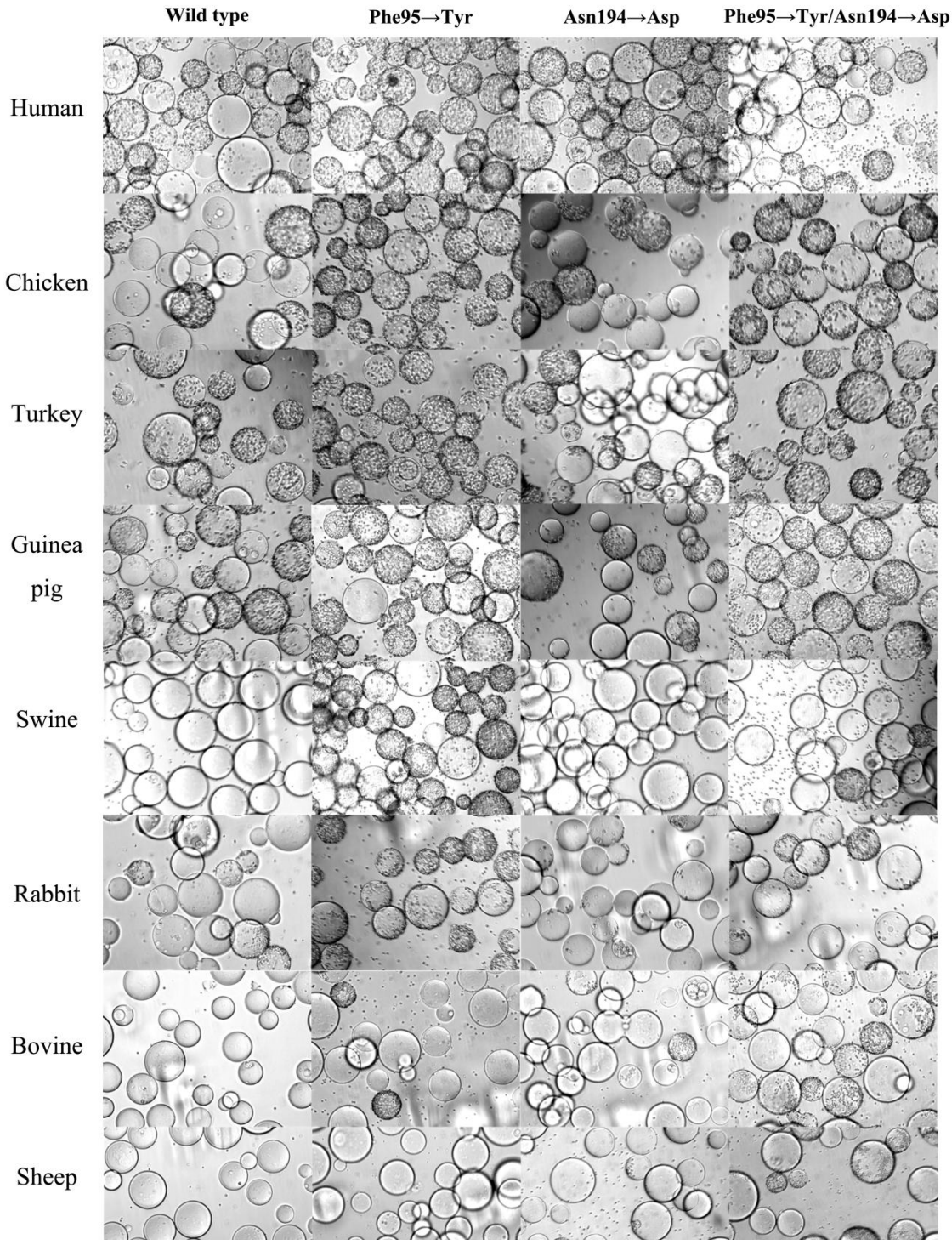

Figure 3.12 Red blood cell (RBC) bindging assay. The species of RBC are listed on the left.

3.2.4. The mutation Phe95→Tyr competitively blocked the binding and infection of influenza A virus

Since binding to cell-surface receptors is a prerequisite for infecting host cells by influenza virus, the key question is whether the much higher binding affinity of the Phe95→Tyr and Phe95→Tyr/Asn194→Asp mutants would allow them to compete more effectively against infection caused by influenza virus. As a stringent test, influenza A/H3N2 virus that generally has a much higher binding affinity for cell-surface receptors than influenza B virus was used [19], an A/Brisbane/10/2007-like isolate TX-419 (H3N2) strain that was kindly provided by Dr. Pedro Piedra. The test also included a "blank" where no influenza A virus was added and a "positive control" where only influenza A virus was added (Figure 3.13). If the binding affinity of BHA and/or its mutants is strong enough to compete against influenza A virus for binding to cultured cells, it is expected to see a lower virus titer in the presence of these recombinant proteins. Madin Darby Canine Kidney (MDCK) was used for the assay. Using five hemagglutination units (HAU) of influenza A virus as input to infect MDCK cells, at 24 hour post-infection, the "positive control" had a virus titer of about 32 HAU per 50 μL 1% human red blood cells. With 400 μg recombinant proteins, the wild type had about two-fold inhibition of influenza A virus over the "positive control" (Figure 3.13), whereas the Phe95→Tyr mutant displayed eight-fold inhibition, representing four-times stronger inhibition. Most strikingly, while the Asn194→Asp single mutant exhibited a very weak inhibition against the binding of influenza A virus, the mutant

Phe95→Tyr/Asn194→Asp completely blocked the infection of influenza A virus, with as low as 100 μg (Figure 3.13) recombinant proteins. In conclusion, the higher binding affinity of the Phe95→Tyr and Phe95→Tyr/Asn194→Asp mutants for synthetic and natural sialic acid receptors allows a much stronger competition against influenza A virus infection.

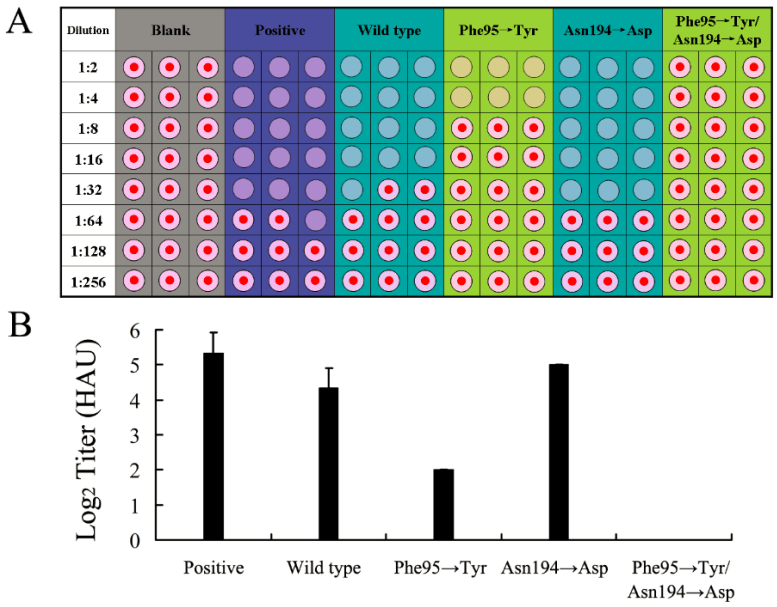

Figure 3.13 The mutation Phe95→Tyr efficiently inhibits the binding of influenza A virus to MDCK cells. The "Blank" sample is the negative control where the MDCK cells grow under normal condition in the binding inhibition assay. The "Positive" sample is where MDCK cells are infected with 5 HAU influenza A virus. The "Wild type" and "Phe95→Tyr" samples represent those wells where the MDCK cells are first exposed to 400 μg wild type BHA or Phe95→Tyr mutant before infection by 5 HAU influenza A viral particles. The "Asn194→Asp" and "Phe95→Tyr/Asn194→Asp" samples are for those wells where the MDCK cells are first bound to 200 μg Asn194→Asp or Phe95→Tyr/Asn194→Asp mutants before the infection of 5 HAU influenza A virus. The supernatants after 24-hour incubation are used for the hemagglutination assay (upper Fig.). If there are enough viruses at a specific dilution, the lattice will form due to the interactions (agglutination) between RBC and hemagglutinin on the surface of influenza A viruses (shown as pink circle); otherwise, the RBC will settle to the bottom of the well to form a red button (shown as a small red circle in a big pink circle). The titer is read as the endpoint of agglutination (lower Figure). The titers are much smaller if Phe95 is mutated to Tyr95 as shown by comparing Phe95→Tyr with wild type or Phe95→Tyr/Asn194→Asp with Asn194→Asp, meaning that the virus could not grow well if Phe95→Tyr or Phe95→Tyr/Asn194→Asp is present. The results indicate that the Phe95→Tyr and Phe95→Tyr/Asn194→Asp mutants could compete against the hemagglutinins on the influenza A virus in the binding to the receptors on the MDCK cells, thus inhibiting the entrance of influenza A virus.

3.2.5. The mutation Phe95→Tyr greatly improved the binding of BHA to three cultured cell lines

Based on the results of the influenza virus infection inhibition assay presented above, the following question was that whether the enhanced receptor-binding affinity of the Phe95→Tyr and Phe95→Tyr/Asn194→Asp mutants allows BHA to bind host cells significantly better. A whole-cell-based ELISA assay was used to quantitatively measure the amount of BHA proteins immobilized on cell culture monolayers due to their association with cell-surface receptors (see Materials and Methods for details). Three established cell lines were used: MDCK, Baby hamster kidney (BHK21), and African green monkey kidney (Vero) (*115*). Compared to the parental MDCK cells, BHK21 cells has a substantially lower level of α(2,6)-linked human-like receptors on the cell surface and supports poorly the growth of influenza A and B viruses (*116*, *117*). Overall, the wild-type BHA exhibited very weak binding to all three cell lines, with the weakest for BHK21 cell line (Figure 3.14A). The mutation Phe95→Tyr substantially improved the binding to all these cell lines, and the most significant improvement was seen for the Vero cell line (Figure 3.14B). The Asn194→Asp mutant had a similar binding affinity as the Phe95→Tyr on the wild-type background (Figure 3.14C), whereas the Phe95→Tyr/Asn194→Asp double mutant substantially enhanced the binding over those of Asn194→Asp (Figure 3.14D). Taken together, the mutation Phe95→Tyr allowed for much tighter binding to all the three cell lines tested here.

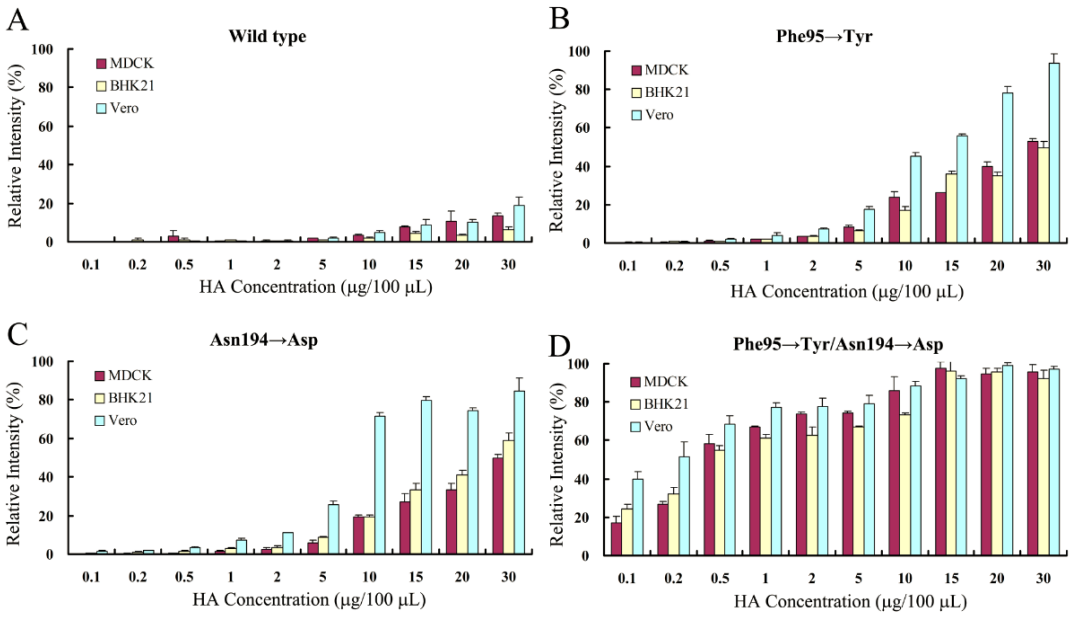

Figure 3.14 The mutation Phe95→Tyr enhances the binding of BHA to the surface of cultured cells. The binding signals are expressed as the percentage of the maximum chemiluminescence intensity for MDCK (A), BHK21 (B) and Vero (C) in cell-based ELISA. In all three cell lines tested, wild-type BHA has the weakest binding, whereas the Phe95→Tyr/Asn194→Asp mutant shows the strongest binding.

3.2.6. Replication of recombinant influenza B viruses in MDCK cells and mice

In the process of infecting hosts, the hemagglutinin protein of influenza virus carries out the following principal functions: binding to cell-surface receptors, signaling endocytosis, and mediating membrane fusion and release of viral genetic materials. In order to ascertain the functions of BHA harboring the introduced Phe95→Tyr mutation, recombinant influenza B viruses were generated by reverse genetics using all genes from influenza B/Lee/40 (generous gifts from Dr. Kawaoka) following previously reported protocols (*103*). The mutations at amino acid sites 95

and 194 in BHA were introduced into the coding region of B/Lee/40 HA by site-directed mutagenesis using Quikchange kit (Agilent). All the viruses have the designed sequences after one passage in MDCK cells. However, in wild-type recombinant B/Lee/40 virus, Asn194 was mutated to Asp194 or Ser194 after the second passage in MDCK cells. Because Asn194 was not stable in MDCK cells in wild-type recombinant B/Lee/40 virus, the studies were only conducted on recombinant B/Lee/40 viruses harboring Asn194→Asp or Phe95→Tyr/Asn194→Asp in BHA.

From the data in Table 3.5, by comparing to the titer of Asn194→Asp virus (at 2.8×10^6 pfu/mL), which can be deemed as an egg-adapted wild-type virus, the Phe95→Tyr/Asn194→Asp virus had a comparable titer of 3.0×10^6 pfu/ml when replicated in MDCK cells. However, the Phe95→Tyr/Asn194→Asp virus was attenuated by ~1,000-fold when it was replicated in mouse lung.

Table 3.5 Replication capabilities of recombinant viruses.

	MDCK[a] (pfu/ml)	Mice lungs[b] (pfu/ml lung)
Asn194→Asp	$2.8\pm0.2\times10^6$	$2.7\pm0.3\times10^5$
Phe95→Tyr/Asn194→Asp	$3.0\pm0.1\times10^6$	$2.2\pm1.7\times10^2$

[a] MDCK cells were infected at MOI of 0.1, and the titer was determined at 48 hours post-infection. [b] Mice were administered viruses intranasally at 10^4 PFU, and lungs were harvested at 3 days postinfection. Titers are expressed as pfu per ml lung supernatant.

3.2.7. Agglutination of erythrocytes by recombinant viruses

The ability of recombinant viruses to agglutinate erythrocytes of different animal species was examined (Table 3.6). Chicken, turkey and guinea pig erythrocytes contain mixtures of α(2,3)- and α(2,6)-linked sialic acid, whereas sheep, horse and bovine erythrocytes contain mainly α(2,3)-linked sialic acid (*118*, *119*). Not only the link type of the sialic acid, but also the density and distribution of the receptor affect the agglutination reaction. The Asn194→Asp virus agglutinates erythrocytes from all species of animals tested but show higher HA titers against erythrocytes bearing both α(2,3)- and α(2,6)-linked sialic acid when the reactions are incubated at 4 degree for 2 h. After further incubation at 37 ℃ for 2h when neuraminidase on the virus begins to cleave the sialic acid bound by BHA, the Asn194→Asp virus could only agglutinate turkey erythrocytes. With incubation at 37 degree overnight, all Asn194→Asp viruses are released from turkey erythrocytes.

Table 3.6 Agglutination of erythrocytes by recombinant viruses under different conditions[a]

Erythrocytes[b]	Agglutination by virus (HAU/50 μl)					
	4 ℃, 2h		4 ℃, 2h and 37 ℃, 2h		4 ℃, 2h and 37 ℃, overnight	
	Asn194→Asp	Phe95→Tyr/Asn194→Asp	Asn194→Asp	Phe95→Tyr/Asn194→Asp	Asn194→Asp	Phe95→Tyr/Asn194→Asp
cRBC	128~256	8~16	<2	8~16	<2	<2
tRBC	256~512	16~32	128~256	16~32	<2	2~4
gpRBC	128~258	16~32	<2	16~32	<2	16~32
sRBC	64~128	64~128	<2	32~64	<2	<2
hoRBC	4~8	64~128	<2	64~128	<2	32~64
bRBC	64~128	16~32	<2	16~32	<2	<2
Asialo RBC	<2	<2	<2	<2	<2	<2

[a]HAU were recorded at different time points after erythrocytes were incubated with virus for 2 h at 4℃, then placed at 37 ℃ for 2 h, and overnight. [b]cRBC, chicken red blood cell (RBC); tRBC, turkey RBC; gpRBC, guinea pig RBC; sRBC, sheep RBC; hoRBC, horse RBC; bRBC, bovine RBC. Asialo chicken RBCs were produced by cleaving sialic acid with *Clostridium perfringens* neuraminidase prior to the agglutination assay.

In contrast, the Phe95→Tyr/Asn194→Asp virus displays relatively higher HA titers against erythrocytes containing α(2,3)-linked sialic acid with the reactions incubated at 4 degree for 2h. After incubation at 37 degree for 2h, the HA titers of the Phe95→Tyr/Asn194→Asp virus against erythrocytes from five species remains same. Even after overnight incubation at 37 degree, this virus still agglutinates turkey, guinea pig and horse erythrocytes with same or lower HA titers.

3.2.8. Cell surface binding of recombinant viruses

The binding of recombinant viruses to the cell surface will directly reflect the ability of the virus to infect the cell. In order to measure this interaction, the fluorescently labeled viruses were added to the pre-chilled cell monolayer. The low temperature will prevent the endocytosis of bound virions. Four different cell lines were used including BHK21, Vero, A549 (human alveolar basal epithelial cells) and MDCK (Figure 3.15). For the Asn194→Asp virus, the amounts of viral particles bound to different cell lines are at the same level as shown by the similar fluorescent signals. For the Phe95→Tyr/Asn194→Asp virus, the signals are about 60% of the signals from the Asn194→Asp virus in BHK21, Vero and A549, and about 80% in MDCK; indicating fewer viral particles bound to the cell surface for this variant.

The NA activity of viral particles at 4 degree were monitored by using substrate MUNANA, and no NA activity was detected. The results of the cell surface binding by recombinant viruses were not affected by the NA activity.

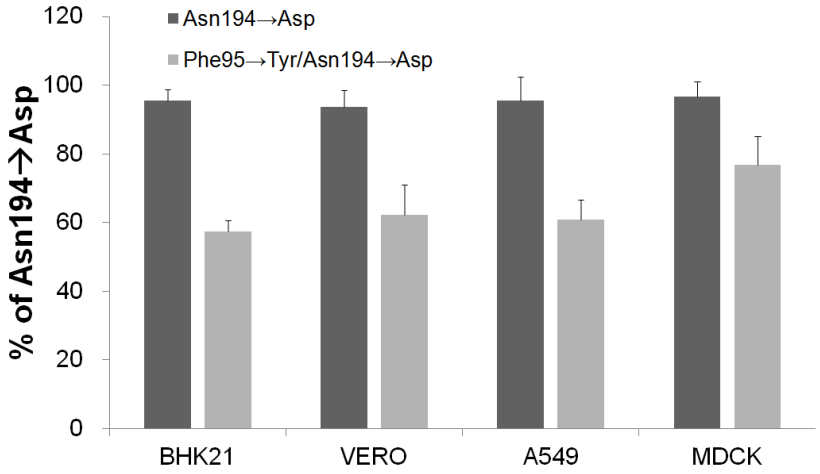

Figure 3.15 Fluorescence binding assays of recombinant viruses. Alexa 488-labeled viruses were bound to chilled cells in 96-well plates at an MOI of 3.0 and incubated at 4 degree for 1 hour. The relative binding for each cell line is expressed as a percentage of Asn194→Asp virus.

3.3. Structure of BHA_2 at the post-fusion state

3.3.1. Data collection and processing

The native dataset suffered from severe radiation damage as judged from the scale factor and B-factor profile from SCALA output. The scale factors increased after 30 frames and the *B*-factor becomes smaller than -5 after 90 frames. After screening for different derivative datasets, one dataset immersed in 1 mM Europium (Eu) was found to have less radiation damage. The collection statistics for this derivative data are shown in Table 3.7. The solution for this dataset was found by molecular replacement, and there is one chain of BHA_2 in the asymmetric unit. The biological trimeric state of BHA_2 could be generated by rotation through the lattice 3-fold symmetric axis.

Table 3.7 Data collection and refinement statistics.

Data collection	
Wavelength (Å)	1.127
Resolution range (Å)	37.76-2.45 (2.54-2.45)
Space group	R32
Unit cell (a b c α β γ)	48.18 48.18 354.895 90 90 120
Total reflections	25553
Unique reflections	5617 (584)
Multiplicity	4.5 (4.4)
Completeness (%)	89.43 (94.96)
Mean I/sigma(I)	15.81 (5.21)
Wilson B-factor ($Å^2$)	40.44
R-sym	0.053 (0.210)
R-factor	0.2549 (0.2733)
R-free	0.2681 (0.3333)
Number of atoms	1111
macromolecules	1084
water	27
Protein residues	142
RMS(bonds)	0.005
RMS(angles)	0.81
Ramachandran favored (%)	91
Ramachandran outliers (%)	2.9
Clashscore	34.07
Average B-factor ($Å^2$)	53.2
Macromolecules	53.4
Solvent	44.4

Statistics for the highest-resolution shell are shown in parentheses.

3.3.2. The overall fold of BHA2

The sequence identity between BHA_2 and AHA_2 is only 29% as calculated by ClustalW2 (*120*). Despite this low sequence identity (Figure 3.16), BHA_2 is still able to fold into a hairpin-like structure (Figure 3.17A), which is very similar to AHA_2 (Figure 3.17B) (*87*, *88*). As for the predicted biological trimer, the common structural features between BHA_2 and AHA_2 include 1) the N-cap domain formed by the N-terminus of the long helix together with the C-terminus of BHA_2, 2) the central coiled-coil assembled from helix A, B and C, 3) the six-helix bundle at the C-terminus

of the long helix and 4) the C-terminal fragment lying between the groove created by central helices.

```
     31         41         51         61         71         81
AHA2 GTGQAADLKS TQAAIDQING KLNRVIEKTN EKFHQIEKEF SEVEGRIQDL EKYVEDTKID
BHA2 GVAVAADLKS TQEAINKITK NLNSLSELEV KNLQRLSGAM DELHNEILEL DEKVDDLRAD
     *.. ****** ** **::*.  :** : *    :::::::  : .*:...* :* :: *:* : *
     91         101        111        121        131        141
AHA2 LWSYNAELLV ALENQHTIDL TDSEMNKLFE KTRRQLRENA EEMGNGSFKI YHKCDNACIE
BHA2 TISSQIELAV LLSNEGIINS EDEHLLALER KLKKMLGPSA VDIGNGCFET KHKCNQTCLD
       * : ** *  *.*:  *:   *..:  * . * :: *  .*  ::***.*:   ***:::*::
     151        161        171        181
AHA2 SIRNGTYDHD VYRDEALNNR FQIKGVELKS GYKDW
BHA2 RIAAGTFNAG EFSLPTFDS- LNITAASLND D----
      *  **:: .  :   :::.  ::*....*:. .
```

Figure 3.16 Protein sequence alignment of AHA_2 and BHA_2. The central helix is shown in red, helix D is in blue, and helix G is in magenta.

Overall, although the D-loop and the C-terminal fragment display some differences between AHA_2 and BHA_2, the coiled-coil regions resemble each other. The amino acid sequence alignment between AHA_2 and BHA_2 is shown in Figure 3.16. Residues 37 to 105 of BHA_2 (AHA_2 numbering) form the central helix, the length of which is same as that of AHA_2. The region between helix D and helix G of BHA_2 lacks of secondary structure, and the density for this region is also weak suggesting great flexibility. For the reported AHA_2 structure where there are two trimers in one asymmetric unit, this region adopts different conformations including α-helix, β-sheet and random coil (*87*). The length of helix G of BHA_2 (5 residues) is shorter than that of AHA_2 (11 residues) (Figure 3.16).

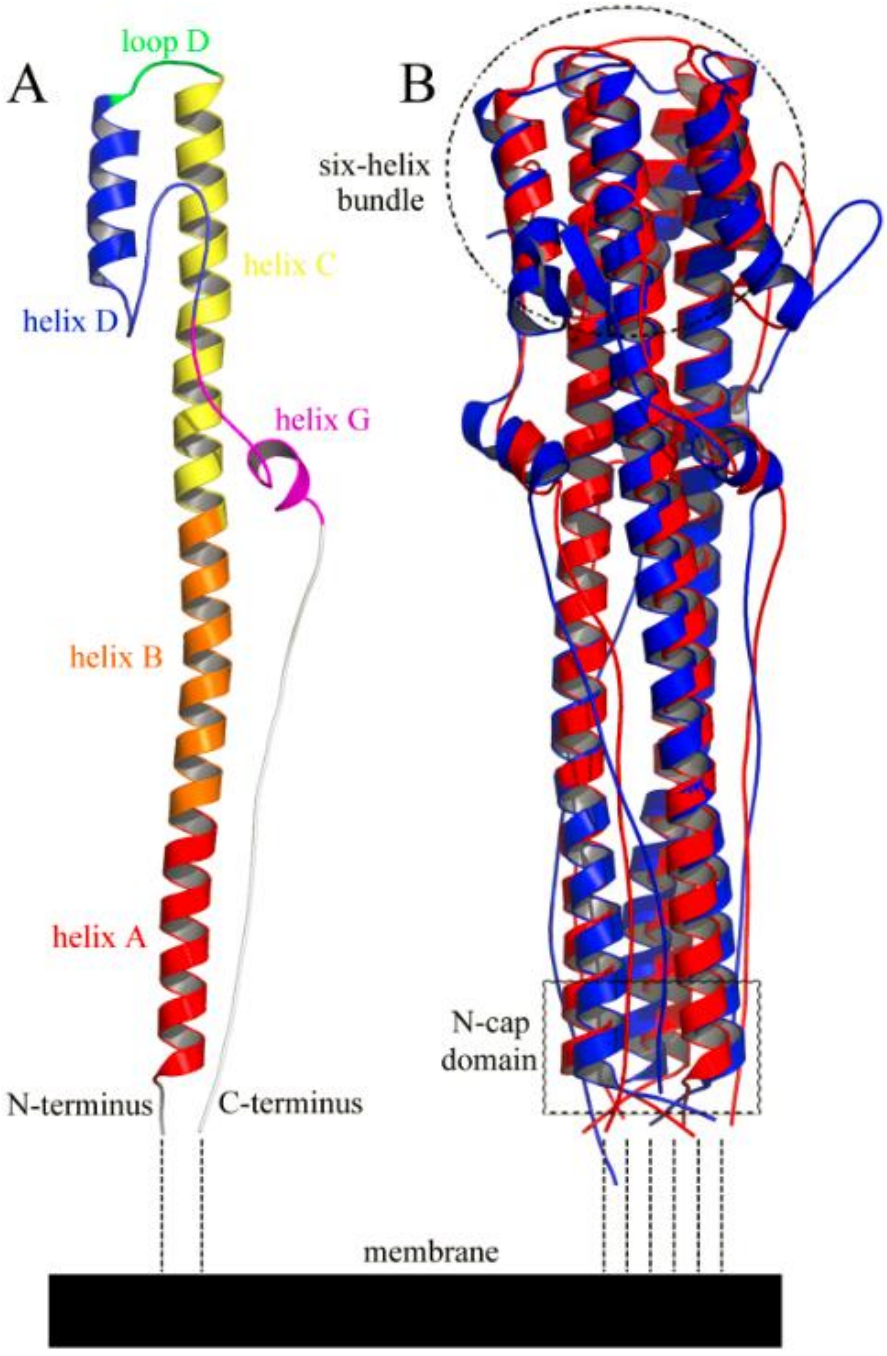

Figure 3.17 Overall fold of BHA_2. (A) The monomer of BHA_2 is colored with structural elements. (B) The trimer of BHA_2 (red) is superposed onto trimeric AHA_2 (blue) (PDB accession code: 1QU1). The fusion peptides at the N-terminus of HA_2 and transmembrane domain at the C-terminus of HA are missing in the structures and are indicated as dash lines. The fusion peptide and transmembrane are brought into close proximity in the membrane (black rectangular) in the post-fusion state.

3.3.3. The N-cap domain

The N-cap domain ties together the N-termini of the coiled-coil and the C-terminal residues. The sequence identities among the N-cap domain (residues 35 to 42) between BHA_2 and AHA_2 are extremely high as seen in Figure 3.16, suggesting similar interaction modes in this domain. Asp37 makes hydrogen bonds by its carbonyl and carboxyl oxygen atoms with Lys39 and Ser40 to stabilize the N-terminal end of the long helix. Ala36 from three monomers forms a nonpolar layer stacking on a layer formed by Ala35 residues.

A subtle difference between BHA_2 and AHA_2 is found in the N-cap domain. The C-terminal fragments of BHA_2 attach to on another less tightly than those of AHA_2. As shown in Figure 3.18B, there are hydrophobic interactions between Val34 and Ala175 in BHA_2, with the distance between two closest atoms of 3.6 Å. As shown in Figure 3.18A, the side-chain of Gln34 in AHA_2 forms hydrogen bonds with carbonyl oxygen atoms of Gly175 and Val176, with the distance of 2.7 Å between the closest interacting atoms. Further, the smaller residue Ala176 of BHA_2 lies farther from the center of the triangle formed by the N-terminal residues than the larger residue Val176 of AHA2 does.

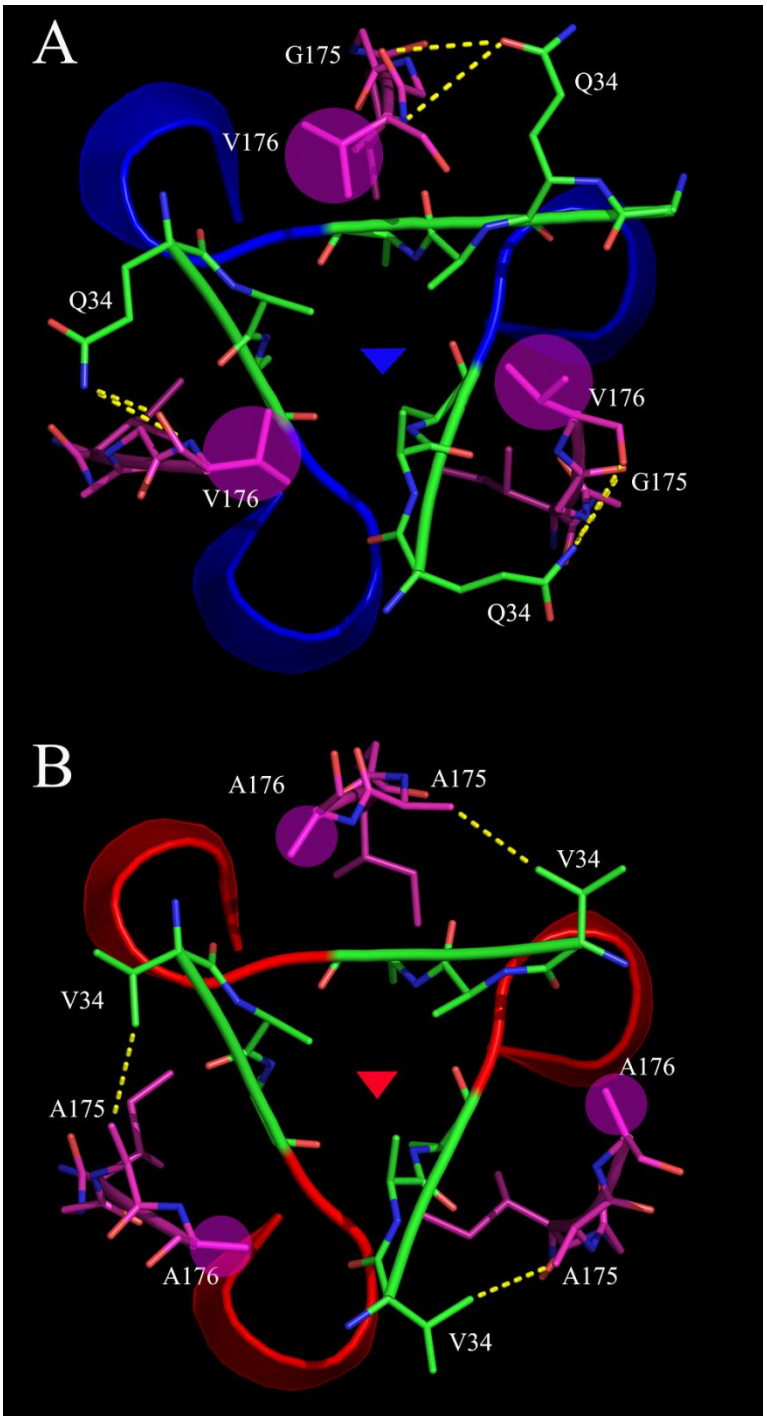

Figure 3.18 Interactions around the N-cap domain of AHA_2 (A) and BHA_2 (B). The first turn of the long helix is shown, and the N- and C-terminal residues are shown in stick mode with green and magenta for carbon atoms, respectively. The three-fold symmetric axis is shown by a small triangle. The side-chains of Val176 of AHA_2 and Ala176 of BHA_2 are highlighted to show their distances from the center of triangle.

3.3.4. The long central coiled-coil

By superimposing the central helix (residue 37 to 105) of BHA_2 on the central helix of the AHA_2 trimer based on their Cα atoms, the RMSD values for each pair of Cα atoms can be calculated as shown in Figure 3.19. Some residues in helix A have small RMSD values, including Thr41, Ile45, Ile48 and Leu52. These four residues are located at the inter-helical interface and are identical between BHA_2 and AHA_2 (Figure 3.16). The other residues in helix A show larger RMSD values, and these residues participate in the interactions with the C-terminal fragment, indicating a different interaction pattern for BHA_2 and AHA_2 between C-terminal fragment and the groove formed by the adjacent helix A.

All residues from 59 to 72 have relatively large RMSD values, and these residues are all in helix B. Helix B is formed from the loop-to-helix transition after HA is exposed to low pH, and this transition largely contributes to the dramatic conformational changes of HA. There is only one identical residue in helix B (Glu72) (Figure 3.16), indicating that BHA_2 employs different residues as compared with AHA_2 not only for interaction with C-terminal fragment, but also for lining up the inter-helical interface around the helix B region.

The residues from 77 to 98 in helix C have much smaller RMSD values in contrast to those in helix B. This sequence also has more similarities (Figure 3.16), and structurally this region is shielded by the outer layer part including helix D to helix G, making it relatively constrained. However, the residues at the end of the helix C (residues 100 to 105) display larger RMSD values, and these residues

interact with helix D to form the six-helix bundle, indicating that BHA_2 must utilize a different strategy to stabilize a six-helix bundle, which is an important intermediate state during fusion.

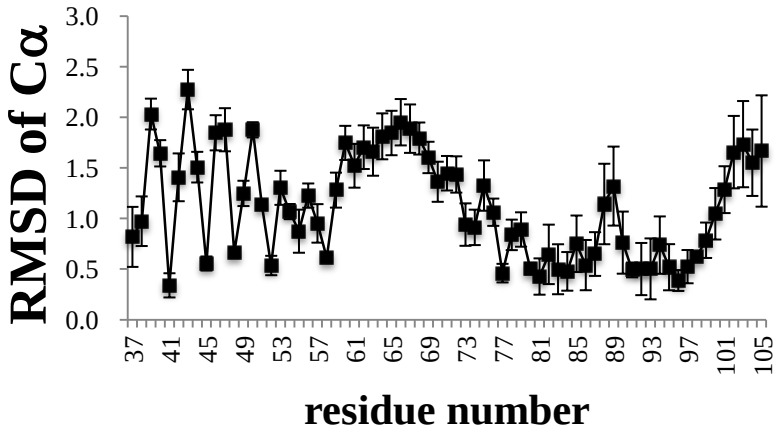

Figure 3.19 RMSD values for Cα atoms of the long central helix between BHA_2 and AHA_2. The BHA_2 monomer is superimposed on each monomer of the AHA_2 trimer by fitting their Cα atoms. The average RMSD values and standard deviations are shown.

3.3.5. The interaction between central coiled coil and C-terminal fragment

The packing of the C-terminal fragment into the groove formed by the central helices will bring the C-terminal transmembrane domain in proximity to the fusion peptide. As illustrated by the RMSD profile, the helices A and B of BHA_2 form the groove where the C-terminal fragment could dock, and BHA_2 would be expected to employ different residues than AHA_2 to stabilize this open hot-dog-like conformation, where the two strands of helices resemble two pieces of bread and

the C-terminal fragment the sausage. To explore the detailed interaction mode in this region, the central helix, including helices A and B, and the C-terminal fragment are shown in detail (Figure 3.20). Overall, it is found that in BHA_2 (Figure 3.20B), the C-terminal fragment is in an extended state, and lies almost in the middle of two helices, making it a parallel conformation with the helices. In contrast, the C-terminal fragment in AHA_2 (Figure 3.20A) reaches across from the C-terminus of one helix to the N-terminus of the other helix; the two helices and the C-terminal fragment form a "z"-like conformation.

The length of the C-terminal fragment of BHA_2 (16 residues) in this region is three-residues shorter than that of AHA_2. The hydrogen bonds in this region were predicted using Pymol and are shown by yellow dashed lines. There are 34 pairs of hydrogen bonds in AHA_2 between the C-terminal fragment and two helices, whereas only 9 pairs are found in BHA_2 (Figure 3.20). In contrast, more hydrophobic residues are found in the C-terminal fragment of BHA_2, and these side-chains face toward the groove. The residues include Phe157, Phe162, Leu164, Pro165, Phe167 and Ile173, as shown in the ball mode in Fig.2.5B; in contrast, AHA2 only has Tyr157, Tyr162 and Ile173 (Figure 3.20A, residues in ball mode) that make contacts within the groove by hydrophobic interaction.

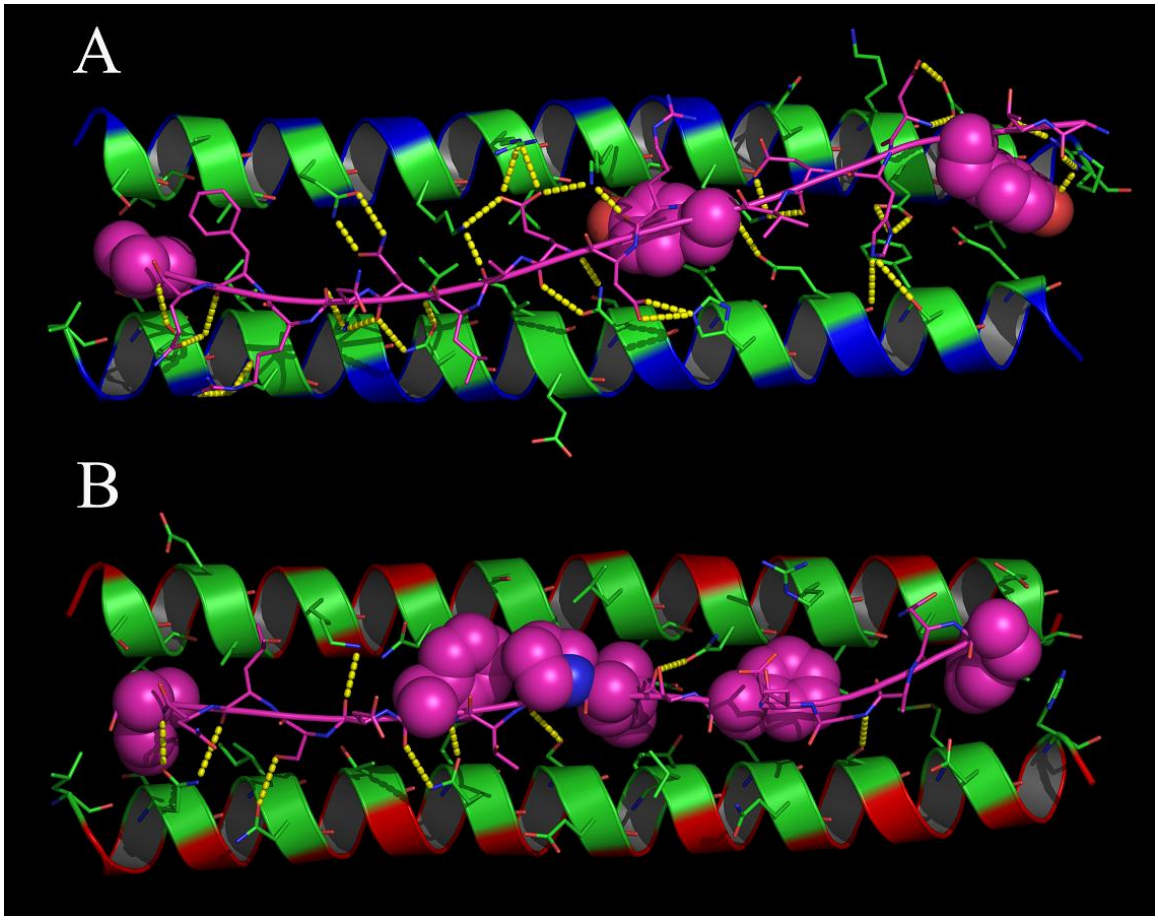

Figure 3.20 Interaction between C-terminal fragment with the groove formed by two adjacent helices (only helix A and helix B are shown) of AHA_2 (A) and BHA_2 (B). The C-terminal fragment is shown in magenta. The residues involved in the interactions are shown in stick mode. Hydrogen bonds shown in yellow dashed lines were predicted by Pymol, and hydrophobic residues of C-terminal fragment facing toward the groove are shown in ball mode.

3.3.6. The six-helix bundle

The six-helix bundle results from a helix-turn transition in a portion of the residues in helix C followed by the inverse packing of helix D against helix C, a characteristic intermediate that is observed for many enveloped viruses (*121*). The formation of this six-helix bundle of influenza virus is required for the formation of

hemi-fusion state (*83*, *121*). To show the relative orientation of the six helices in the bundle, the structure is rotated to be viewed along the three-fold axis (Figure 3.21). In BHA_2 (Figure 3.21B), Ile107 closely packs along the symmetry axis to form a tight hydrophobic core, surrounded by another layer provided by the hydrophobic residue Ile108. In contrast, in BHA_2, the carboxyl oxygen atoms of Glu105 at the end of helix C interact with the backbone atoms of Asn109 and Ser110 in the loop D of the adjacent chain, stabilizing the arrangement by multiple electrostatic interactions. The N-terminus of the helix D is located in the middle of two neighboring helix C in BHA_2, as illustrated by two inversely placed regular triangles (Figure 3.21B).

As compared with AHA_2 (Figure 3.21A), Thr107 residues form an annulus that covers the C-terminus of the coiled-coil. The hydroxyl group of Thr107 is also close to the carbonyl oxygen atom of Thr107 from the other chains. Moreover, two hydrophobic residues Ile108 and Leu110 of AHA_2 stack mainly between helix C and D, which pushes away the N-terminus of the helix D from helix C.

The stacking orientation of the helix D against the two central helices C in BHA_2 (Figure 3.21B, helix D is in warm pink) is also different from that in AHA_2 ((Figure 3.21A, helix D is in sky-blue). The helix D of BHA_2 itself is straight, while that of AHA_2 is a little bent.

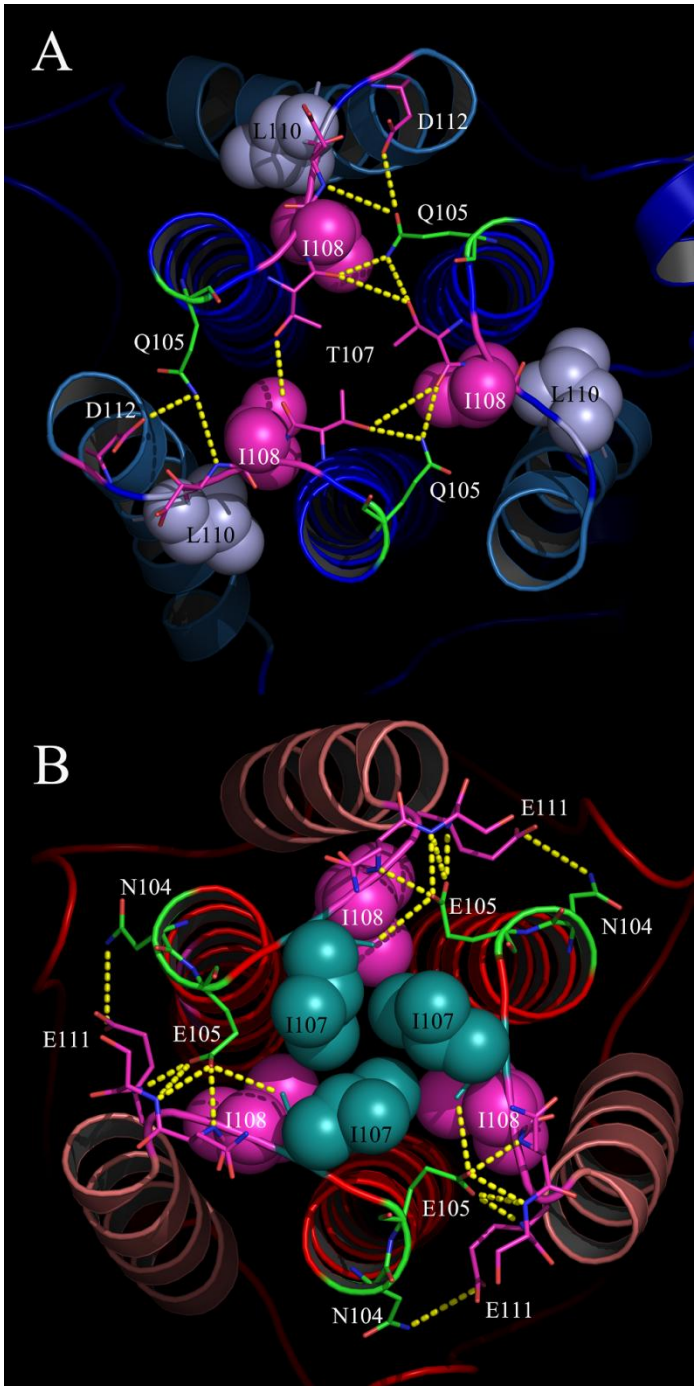

Figure 3.21 Interaction between the loop D and helices in six helix bundle of AHA_2 (A) and BHA_2(B). The central helices of coiled-coil are shown in blue in AHA_2 and in red in BHA_2. The helices D are showin skyblue in AHA_2 and in warmpink in BHA_2.

To further characterize how helix D packs onto helix C in the six-helix bundle, the central coiled-coil is represented in surface mode, and one helix D is shown above the coiled-coil in Figure 3.22. In BHA_2 (Figure 3.22B), no larger residues are found to line the groove formed by two adjacent helix C segments, instead there is a cavity near Leu98. Leu118 in helix D of BHA_2 is just located above the cavity and covers this region along with Leu115, Leu122 and Met125. As compared with AHA_2 (Figure 3.22A), the surface formed between two neighboring helix C segments is rather flat due to the existence of two large hydrophobic residues Trp92 and Tyr94 (BHA_2 has Ile92 and Ser94 at the corresponding sites).

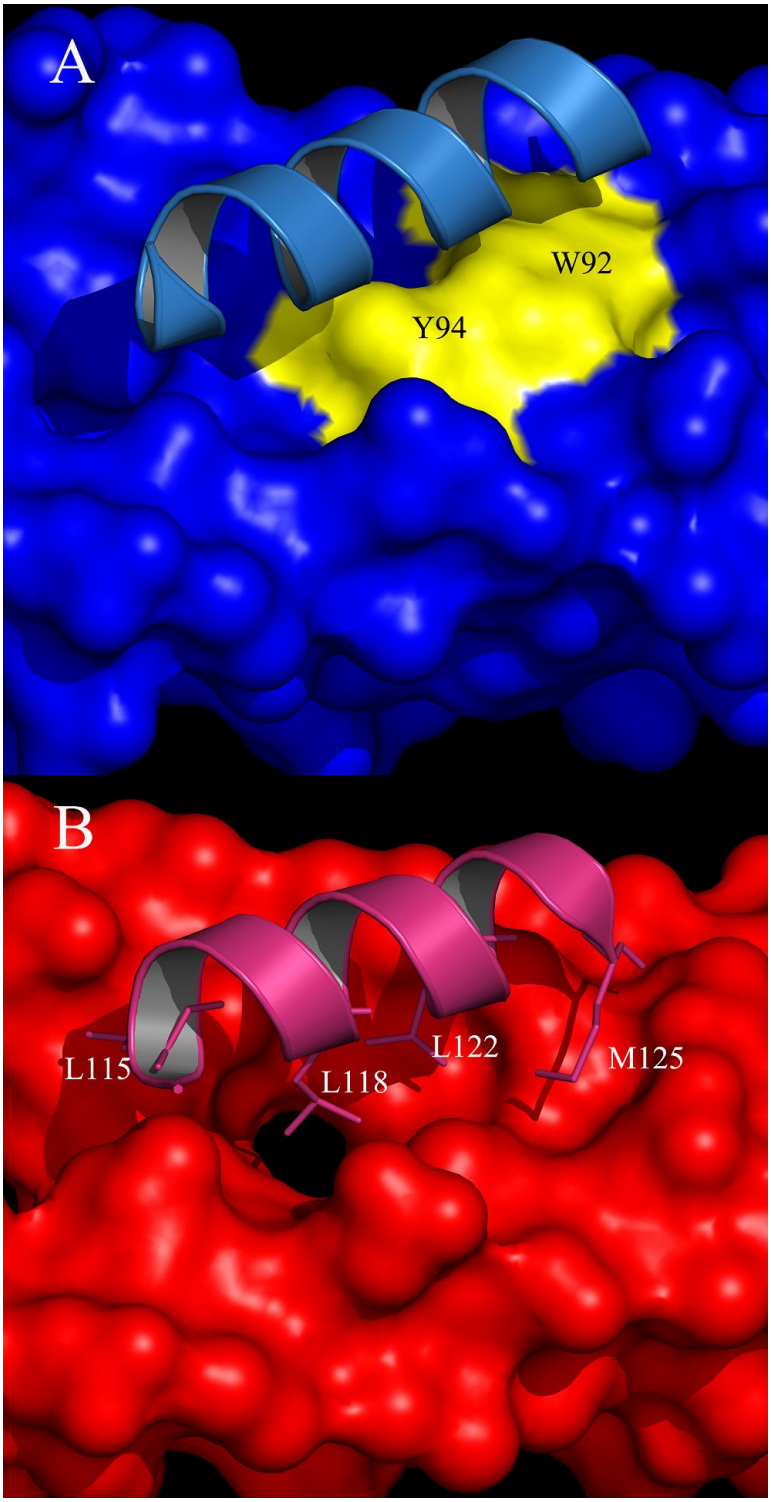

Figure 3.22 Packing of helix D on the neighboring helices C of AHA_2 (A) and BHA_2 (B) in the six-helix bundle. The central coiled-coil is represented by surface mode. The helix C is shown in cartoon mode.

Chapter 4

Discussion

4.1. Normal mode refinement improve the structural quality

Structural refinement at limited resolutions is generally difficult and very inefficient. Here it is presented the application of a new normal-mode-based X-ray crystallographic refinement method (NMRef) to a total of eight low-resolution crystal structures (in the range of 3.0~3.9 Å). All these structures have low data-to-parameter ratio (ranging from far below to barely 1.0). The first two systems (PDB accession codes: 2A0L and 2BBJ) have very large average *B*-factors, indicating a high overall structural flexibility. Two other systems (PDB accession codes: 2PM9 and 1SUQ) contain structural regions with high *B*-factors. Possibly due to the limited data-to-parameter ratio, three systems (PDB accession codes: 2AEQ, 2INY and 2V1C) have a constant B-factor for all atoms, and one system (PDB accession code: 2BBJ)

has an almost constant *B*-factor for all atoms. Of the eight structures, three are membrane proteins, and five are soluble supramolecular complexes.

In all cases, the application of NMRef substantially improved the fit between the structural models and the diffraction data as reflected in the decreases of R-factors (Figure 4.1). The decreases are in the range of 1.7% ~ 7.7% for R_{free}-factors and 1.5% ~ 14.8% for R_{cryst}-factors. The averaged decreases are 3.6% for R_{free}-factor and 4.8% for R_{cryst}-factor. In addition, in most cases, there is a concomitant improvement in the geometry of the structural model, as reflected in the improvement of residue distribution in the Ramachandran plot and reduction in the number of bad contacts. It is worth emphasizing that these substantial improvements were achieved in only one round of normal-mode-based refinement without any manual adjustment, with 5~100 fold fewer thermal parameters than in the original isotropic *B*-factor refinement. As demonstrated in the previous studies, the use of multiple rounds of normal-mode refinement generally brings about much more substantial improvements (*50*).

Since TLS refinement uses a different concept, whether the combined use of TLS with normal-mode refinement will result in further improvement is tested. Indeed, in about half cases, the application of TLS to normal-mode refined structures further decreased the *R*-factors and sometimes improved the geometry as well. The most significant advantage of NMRef over TLS is seen in the last three systems in Figure 4.1. In these systems, the isotropic *B*-factors were set as a constant in the final structures deposited in PDB, which made the direct application

of multi-group TLS refinement impossible. The application of NMRef is not at all affected by the values of isotropic *B*-factors. However, in certain cases, the TLS refinement may deliver better *R*-factor improvement than NMRef. This result arises from the independence of body motion in TLS that may offer a greater flexibility for fitting in cases where structural deformations are less collective.

As clearly indicated from the study on these test systems, testing the normal-mode refinement using a different number of lowest-frequency normal modes is a useful approach. In real applications, testing the combined use of normal-mode refinement with TLS, which in some cases can result in further improvement in structural refinement, is also useful. Even for the cases where TLS outperforms NMRef, an initial application of NMRef may improve the subsequent application of TLS because the former can deliver a smoother *B*-factor distribution for better grouping of TLS.

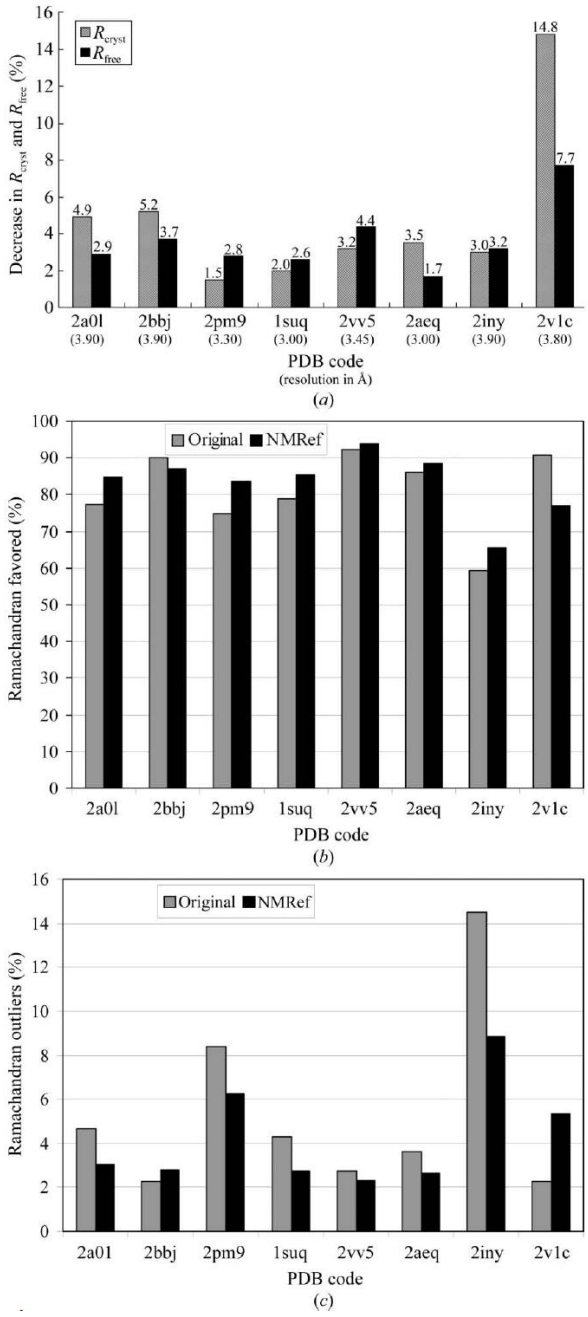

Figure 4.1 Structural improvements produced by normal-mode refinement over the original models. (a) Decreases in *R* factors. Grey bars represent decreases in R_{cryst} and black bars those in R_{free}. (b) Percentage of residues distributed in the Ramachandran favored region and (c) percentage of residues distributed in the Ramachandran outlier region, as calculated using MolProbity (*114*). Grey bars are for the original models and black bars are for the normal-mode refined models.

4.2. Phe95→Tyr increases the receptor-binding affinity of influenza BHA

Compared with influenza A virus, naturally occurring influenza B virus has a much lower binding affinity for synthetic glycans and natural receptors on the cell surface (*94*). Previous structural studies of influenza BHA indicated a critical difference between the HA glycoproteins of influenza A and B viruses at the bottom of the receptor binding sites, Tyr98 in AHA versus Phe95 in BHA (*69*, *70*). Given the absolutely conserved nature of Tyr98 among all known 16 subtypes of influenza AHA proteins, the presence of Phe95 in BHA might be responsible for the lower receptor binding affinity of influenza B virus (Figure 1.4).

By using recombinant influenza BHA and its designed variants, the results demonstrate that Phe95→Tyr is able to enhance the binding affinity of BHA to synthetic glycans significantly (Figure 3.10, Table 3.4), corresponding to a level comparable to those of influenza A virus (*64*, *122–124*) and allowed BHA to bind red blood cells from many different animals (Figure 3.12) that are known to contain a great deal of variation in cell-surface glycan compositions. Most significantly, the "corrective" mutation can competitively block the binding and infection of influenza A virus (Figure 3.13) and greatly improve the binding of mutant BHA to all three cell lines that showed very low binding to wild-type BHA (Figure 3.14).

An outstanding feature of the improved receptor binding affinity by the "corrective" mutation, Phe95→Tyr, is the almost indiscriminate enhancement

towards both human-like and avian-like receptors (Figure 3.10, Table 3.4). This binding improvement is postulated to be the fact that the Sia-1 moiety interacts with the hydroxyl group of Tyr98 of H3 AHA via two hydrogen bonds (Figure 1.4A), a pattern shared by both types of receptors. Thus, it is expected the introduced hydroxyl group of Tyr95 in BHA to make similar interactions with the Sia-1 moiety. To confirm that this arrangement is structurally feasible, this mutation in the BHA crystal structure is made *in silico* (PDB accession code: 2RFU) (*70*). With only minor adjustment, the side-chain of Tyr95 was readily placed to a position similar to that of Tyr98 in H3 AHA, and similarly makes two hydrogen bonds with the Sia-1 moiety and one hydrogen bond with His191, another key residue of the RBS (Figure 1.4C). This hypothesis was further supported by a glycan array analysis in which Phe95→Tyr and Phe95→Tyr/Asn194→Asp mutants were able to bind significantly better to the glycan containing only the Sia-1 moiety than wild-type and Asn194→Asp BHAs, respectively (Figure 3.10) (*125*).

In the binding competition assay (Figure 3.13), the ability of BHA or its various mutants to inhibit the binding of influenza A virus appeared to be closely related to the receptor binding affinity of corresponding BHA glycoproteins to α(2,6)-linked receptors (Table 3.4). For instance, the slightly better binding of wild-type BHA, compared to that of Asn194→Asp, corresponded to slightly stronger inhibition against influenza A virus. Similarly, the 10-fold higher binding affinity of Phe95→Tyr/Asn194→Asp over that of Phe95→Tyr resulted in at least four-fold of stronger inhibition. This result is presumably because the influenza A virus that

used in this assay, an A/Brisbane/10/2007-like strain, is a human strain that would preferentially bind to α(2,6)-linked sialic acid receptors. Thus, the better the binding affinity of BHA to α(2,6)-linked sialic acid receptors, the fewer the α(2,6)-linked sialic acid receptors remain accessible for A/Brisbane/ 10/2007-like viruses to attach and invade the cells and, consequently, lower the viral titers. Moreover, although MDCK cells have both α(2,6)-linked and α(2,3)-linked sialic acid receptors, the former is thought to be in a much higher abundance than the latter (*126*).

Due to the complexity of sialic acid presentations on cell surface, a quantitative calculation of apparent dissociation constants between BHA and three cultured cell lines was not possible. However, if the relative intensity of wild type is compared with the Phe95→Tyr mutant at the highest BHA concentration (both were not saturated), about 5~10 fold better binding for Phe95→Tyr mutant (Figure 3.14B) over the wild type (Figure 3.14A) is seen. Furthermore, if comparing the concentration at which 50% relative intensity was achieved for Asn194→Asp (Figure 3.14C) and Phe95→Tyr/Asn194→Asp (Figure 3.14D) (since the saturation was observed at higher Phe95→Tyr/Asn194→Asp concentration), binding for the protein with the Phe95→Tyr mutation is 30~60 fold better. Therefore, the Phe95→Tyr mutation significantly promotes the binding of BHA to the surface of all three tested cultured cell lines.

Interestingly, the enhanced signal by the side-chain of the mutation of Phe95→Tyr in this study can be accommodated in the receptor binding site of BHA easily and moreover greatly enhances the receptor binding affinity. In sharp

contrast, the large-to-small mutation of Tyr98→Phe on H3 AHA almost abolished the binding of influenza A virus to red blood cells, and the recombinant Tyr98→Phe virus cannot infect mutant MDCK cells with a reduced level of cell-surface sialic acids (*127*). Since Tyr98 contributes one hydrogen bond to stabilize the RBS and two additional ones to bind to the sialic acid receptors (*128*), it is likely that Tyr98→Phe somehow distorts the RBS and/or substantially weakens the receptor binding in H3 AHA. Nevertheless, the fact that wild-type BHA can still bind to sialic acid receptors and red blood cells, albeit at a lower affinity, suggests that the loss of the hydroxyl group in Phe95 has been at least partially compensated, perhaps by some small local structural adjustments such as those found within the 240-loop and 140-loop (*70*).

4.3. The roles of receptor binding of HA in influenza virus pathogenicity

Recent studies on influenza A virus have demonstrated that manipulating receptor-binding avidity is an important tool that influenza A virus can use in response to neutralization imposed by neutralizing antibodies, which seems to drive antigenic drift (*129*). Moreover, a single base-pair change in 2009 H1N1 AHA increased its binding affinity for human receptors, corrected the low receptor affinity of the field isolate, and led to efficient airborne transmission in ferrets (*122*). These and other studies (*123*, *124*, *130*) suggest that the binding affinity of AHA to receptors is an important determinant for influenza infectivity.

The lack of reservoirs of influenza B virus outside of human population is believed to be one of the major reasons that influenza B virus does not undergo antigenic shift, a molecular evolutionary mechanism that is responsible for all known human pandemics caused by influenza A virus (*71*, *128*, *131*). It remains an open question whether there exists a correlation between the limited host range and the lower receptor binding affinity of BHA of natural influenza B virus. Most recently, Palese and co-workers (*132*) published an interesting study in which the ectodomain of influenza PR8 H1 AHA was used to replace that of influenza BHA in recombinant influenza B virus. In marked contrast to the wild-type influenza B virus, the recombinant influenza B virus with inserted PR8 H1 AHA resulted in significant weight loss in infected mice, suggesting that PR8 H1 AHA has somehow enhanced the pathogenicity of the recombinant virus in which it resides.

Taken together, the substantially enhanced receptor binding affinity of BHA upon a single Phe95→Tyr mutation at the RBS site provides a plausible explanation for the very limited host range displayed by influenza B virus. As suggested in previous studies e.g., (*127*), a sufficient density of cell-surface sialic acids with a weaker receptor-binding HA or a lower than optimal density of sialic acids with a stronger receptor-binding HA are both viable combinations for effective influenza infection. Thus, the weaker receptor binding of wild-type BHA imposes a much higher demand for the "perfect" glycan composition on the surface of target cells in a given host; thus, influenza B virus only infects hosts that satisfy such strict requirement. A single Phe95→Tyr mutation was able to promote the binding affinity of BHA to a comparable level as that of AHA, regardless of human or avian-

like receptors, thus enabling BHA to bind a much wider range of red blood cells and three cultured cell lines where the wild-type BHA binds very poorly. Although it is not known when, how and why influenza BHA acquired Tyr95→Phe mutation in the course of evolution, it is likely that some or all of the other 10 viral proteins of influenza B virus have undergone concomitant changes to maintain a delicate balance between BHA and neuraminidase (*133*). Although the data revealed the importance of a single-residue mutation at the receptor binding site to the receptor binding affinity of BHA, further studies are needed to explore implications for pathogenicity and evolution of influenza B virus.

4.4. The increased receptor binding ability of BHA alone may not be advantageous to the replication of viruses

Our studies based on the recombinant proteins show that the Phe95→Tyr mutation increased the receptor binding ability of BHA toward synthetic and natural receptors at protein level. However, the replication of the Phe95→Tyr/Asn194→Asp virus in mice is not as robust as that of the Asn194→Asp virus (Table 3.5). The HA titer of the Phe95→Tyr/Asn194→Asp virus against chicken, turkey and guinea pig erythrocytes is much lower than that of the Asn194→Asp virus when the reactions are incubated at 4 degree for 2 h (Table 3.6). Moreover, the fluorescently labeled virus binding assay indicates that there are about 20%~40% fewer Phe95→Tyr/Asn194→Asp viral particles bound to the four different cell lines (Figure 3.15). Nevertheless, Phe95→Tyr/Asn194→Asp viruses

are still able to agglutinate erythrocytes of three species after incubating at 37 degree for overnight whereas Asn194→Asp completely loses this ability (Table 3.6). Two possible explanations for this observation include: 1) the mutated BHA binds to a receptor that neuraminidase could not recognize, or 2) the mutated BHA binds to the receptor so tightly that access of neuraminidase to the receptor is less efficient. As shown by previous studies, the balance between HA and NA is important for the replication of influenza virus(*134*, *135*). The attenuated replication of the Phe95→Tyr/Asn194→Asp virus in mice may be a result of imbalance between the high receptor binding affinity of BHA with other components of influenza B virus that have changed to accommodate the low BHA binding affinity. Further experiments are needed to test these possibilities.

4.5. Structure of BHA_2

The influenza virus HA protein is a member of class I viral fusion proteins, which are characterized by trimeric hairpin structures containing a central coiled-coil. Human immunodeficiency virus 1 gp41, Ebola virus gp2 and human Parainfluenza virus F also belong to this class of viral fusion proteins (*82*, *136*). Studies on protein-mediated membrane fusion by these proteins not only reveal the mechanisms by which viruses merge with host membranes, but also identify those principles guiding the communication between cell compartments, including, for example, plasma-membrane repair and endosome-lysosome fusion (*85*, *137*, *138*).

The studies on the viral fusion proteins will contribute to the explanation of the evolutionary principles of membrane fusion (*84*).

Early knowledge regarding membrane fusion came largely from studies of influenza A virus HA, and the spring-loaded mechanism was proposed based on the finding that the synthetic 36-residue long peptide comprising the B-loop of AHA_2 folded into a three-stranded coiled coil in solution (*89*). The neutral pH form of AHA in the pre-fusion state was thought to be meta-stable, and the energy stored in this state was released to induce hemifusion triggered by low pH (*121*, *139*, *140*). At a later stage, the formation of the hairpin-like conformation was crucial to the opening and enlargement of the fusion pore (*121*, *141*).

Although class I viral fusion proteins from different viruses are unrelated in protein sequence, these proteins show some common structural features, including a six-helix bundle (*136*). Whether BHA_2 would fold into a hairpin conformation similar to AHA_2 or is precluded from doing so due to the lack of structure of BHA in the post-fusion state remains unknown. Analysis of the structure of BHA_2 together with the reported AHA_2 structure demonstrates some unique interaction modes in those structural elements common for class I viral fusion proteins, and this analysis provides detailed information on how divergent protein sequences could evolve to share the spring-loaded mechanism for fusion.

Importantly, the B-loop of AHA_2 has two large residues Phe63 and Phe70, which are thought to be the key residues for the formation of newly formed coiled coil (*90*, *142*), whereas BHA_2 has Leu63 and Met70 at corresponding sites. BHA_2 and

AHA_2 have extreme sequence diversity in the B-loop (Figure 3.16) and they also show large structural differences (Figure 3.19). Considering the important role of the B-loop in the conformational change, the sequence difference might be interpreted to signal the divergence of BHA and AHA structure and function. On the contrary, the B-loop of BHA_2 has higher probability (Figure 4.2) to form coiled-coil structure than that of AHA_2 calculated by the COILS/PCOILS webserver (http://toolkit.tuebingen.mpg.de/pcoils) (*143*). This result indicates that the B-loop of BHA_2 may form a coiled-coil despite the fact that it lacks the two Phe residues. However, this protein may have a different folding pathway to form the coiled-coil as compared with AHA_2.

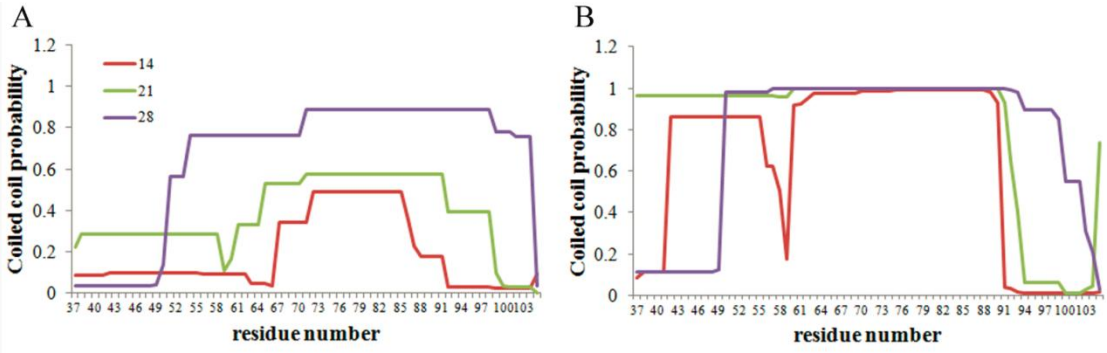

Figure 4.2 The coiled coil probability of AHA_2 (A) and BHA_2 (B) predicted by the COILS/PCOILS server. The prediction is calculated with different lengths of residues as window size (14, 21 or 28 residues) shown in different colors. The B-loop of AHA_2 includes residues 56 to 75 and the B-loop of BHA_2 includes 58 to 74.

His74 is located at the C-terminal end of the B-loop in BHA. The loop conformation around His74 is stabilized by its interaction with Glu76 on the neighboring chain (Figure 4.3) as observed in the pre-fusion BHA structure (*69*). In

group 2 AHA, a conserved His106 is found at the N-terminus of the D-loop; whereas a conserved Arg106 is in group 1 AHA. When an Arg106→His mutation is introduced to H2 AHA (group 1), this side-chain interacts with Glu105 on the neighboring chain in the pre-fusion state to retain the helical conformation around His106 even under acidic condition. The results for Arg106→His in H2 AHA lead to an inference that the stabilization of this residue of HA_2 in the pre-fusion state might affect the fusion ability HA. Histidine residues are often considered as triggers of conformational changes, the so-called "histidine switch" (*144*, *145*). To establish whether His74 of BHA_2 is key to the fusion ability of influenza virus will require further studies.

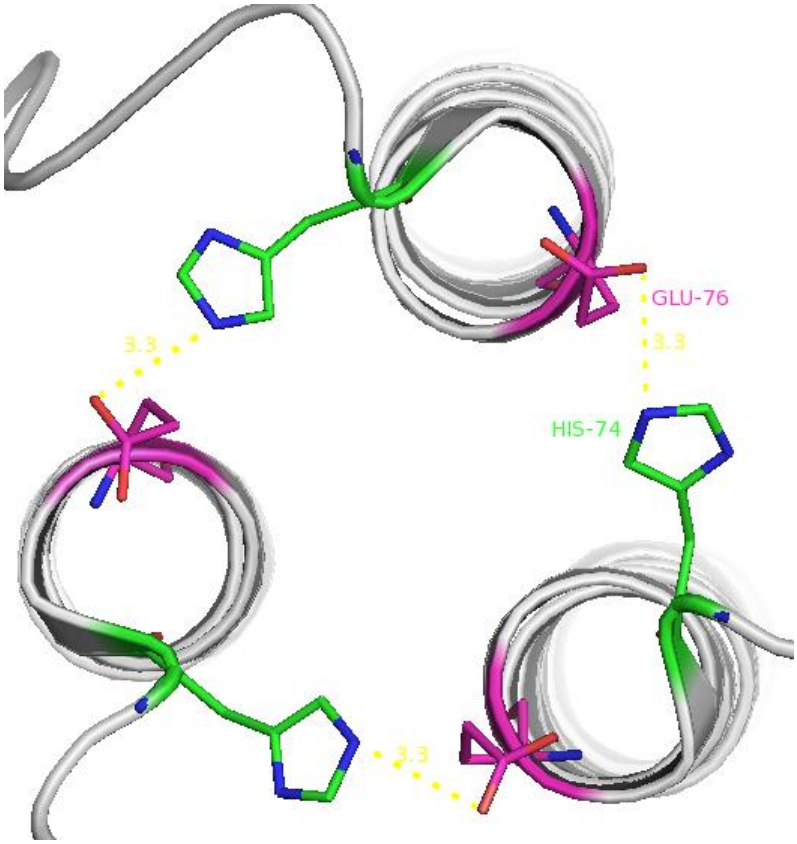

Figure 4.3 His74 located at the C-terminal end of B-loop of BHA forms hydrogen bond with Glu76 of neighboring chain (PDB accession code: 3BT6).

The side-chain of Glu59 in the B-loop of BHA_2 points toward the three-fold axis (Figure 4.4). A buried glutamate residue is found to be responsible for the conformational change induced by low pH in synthetic peptides based on macrophage scavenger receptors (*146*). In the study of those synthetic peptides (*146*), the peptides adopted a random coil structure at neutral pH, and the peptides became an α-helical coiled-coil at pH 5. The glutamate in the synthetic peptides was found to be crucial for the pH-induced conformational change. For BHA at the neutral pH form (*69*), Glu59 is also located in a loop region (N-terminus of B-loop), it is reasonable to assume that this residue serves as the pH sensor when BHA is protonated in the low pH environment of endosome. The protonation of Glu59 will reduces the electrostatic repulsion of this charged residue from each other and facilitates the burial of Glu59 in the interior of the coiled-coil. However, BHA_2 alone (without BHA_1) is still able to form hairpin conformation at the neutral pH. The unanswered question is that how the charged Glu59 is stabilized in the BHA_2 structure. In the BHA_2 structure (Figure 4.4), Glu59 is surrounded by residues, including Asn52, Leu53 and Leu164, that may help the burial of the charged residue in the hydrophobic core.

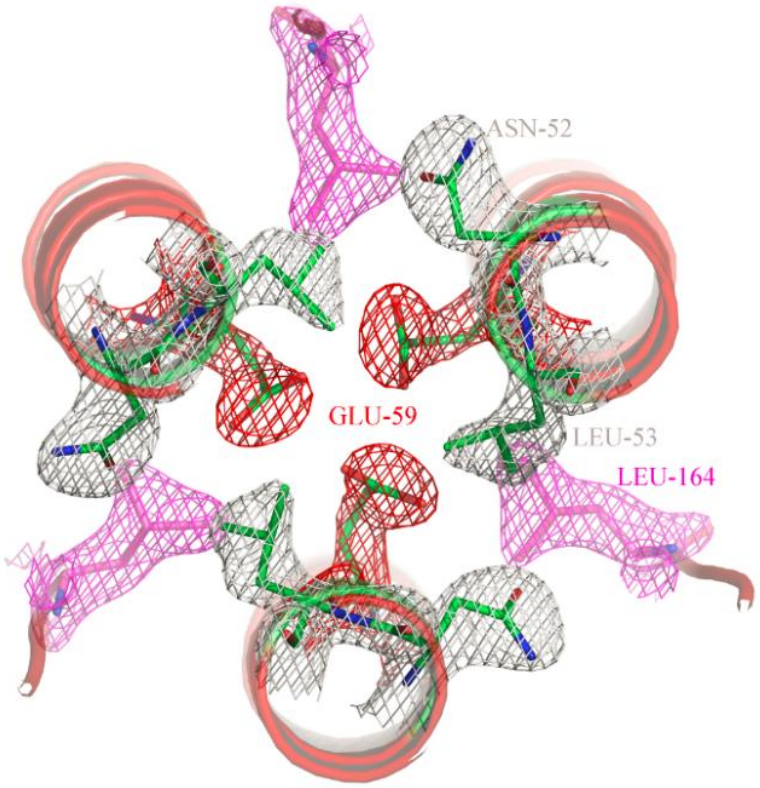

Figure 4.4 Glu59 in the B loop of BHA_2 faces inward and is stacked by the surrounding large residues. The σ-weighted 2FOFC map is shown at 2σ level.

Chapter 5

Conclusions

Influenza B virus remains a health threat to people all over the world. The studies on the structural and functional properties of influenza B virus HA (BHA) will further the understanding of this virus. The normal mode refinement (NMRef) is a useful method to improve the fit between the structural model and diffraction data especially for the structures determined at low resolution and the structures with high flexibilities. By applying the NMRef method, quality of the structural model is improved and thereby facilitating the more accurate interpretation of the interaction patterns in the structure. The usefulness of the NMRef method will possibly be demonstrated on the BHA-antibody crystal structures in the future.

The introduction of Tyr at site 95 in BHA protein could increase the receptor binding affinity of BHA at the protein level. However, recombinant viruses with Phe95→Tyr mutation in BHA do not clearly display the increased receptor binding affinity. The imbalance between the high receptor binding affinity of BHA and un-

modified receptor cleavage ability of influenza B virus NA may affect the pathogenesis of influenza B virus; thus, affecting the receptor binding affinity of the virus. However, how this imbalance affects the pathogenesis of influenza B virus requires further investigations. This study also illustrates the importance and necessity of the co-evolution of HA and NA of influenza viruses.

The structure of BHA_2 at the post-fusion state presents the detailed interaction modes that are involved in the stabilization of the hairpin-like conformation of BHA_2. By comparing the BHA_2 model with the reported AHA_2 structure (*87*), it shows how diverse protein sequences between AHA_2 and BHA_2 may fold into similar conformation. The structure of BHA_2 could be the basis for further investigation of the fusion mechanism for influenza B virus.

Notes

1. Results for the normal mode refinement on eight systems have been published in *Acta Crystallography. Section D. Biological Crystallography* (2009), 65, 633-643 .

References

1. P. Palese, M. Shaw, in *Fields Virology, 5th Edition*, (2007), pp. 1648–90.

2. J. K. Taubenberger, D. M. Morens, The pathology of influenza virus infections., *Annual Review of Pathology* **3**, 499–522 (2008).

3. Y. Suzuki, M. Nei, Origin and evolution of influenza virus hemagglutinin genes., *Molecular Biology and Evolution* **19**, 501–9 (2002).

4. P. Wright, G. Neumann, Y. Kawaoka, in *Fields Virology, 5th Edition*, (2007), pp. 1693–741.

5. R. G. Webster, W. J. Bean, O. T. Gorman, T. M. Chambers, Y. Kawaoka, Evolution and ecology of influenza A viruses., *Microbiological Reviews* **56**, 152–79 (1992).

6. J. K. Taubenberger, J. C. Kash, Influenza virus evolution, host adaptation, and pandemic formation., *Cell Host & Microbe* **7**, 440–51 (2010).

7. A. M. D. E. Osterhaus, G. F. Rimmelzwaan, B. E. E. Martina, T. M. Bestebroer, R. A. M. Fouchier, Influenza B virus in seals., *Science* **288**, 1051–3 (2000).

8. K. J. Jakeman, M. Tisdale, S. Russell, A. Leone, C. Sweet, Efficacy of 2’-deoxy-2'-fluororibosides against influenza A and B viruses in ferrets., *Antimicrobial Agents and Chemotherapy* **38**, 1864–7 (1994).

9. M. C. Zambon, Epidemiology and pathogenesis of influenza., *The Journal of Antimicrobial Chemotherapy* **44**, 3–9 (1999).

10. S. J. Anthony *et al.*, Emergence of fatal avian influenza in New England harbor seals., *mBio* **3**, e00166–12 (2012).

11. D. J. D. Earn, J. Dushoff, S. A. Levin, Ecology and evolution of the flu., *Trends in Ecology & Evolution* **17**, 334–40 (2002).

12. R. G. Webster, M. T. Berton, Analysis of antigenic drift in the haemagglutinin molecule of influenza B virus with monoclonal antibodies., *The Journal of General Virology* **54**, 243–51 (1981).

13. A. Burke, M. D. Cunha, Influenza: historical aspects of epidemics and pandemics., *Infectious Disease Clinics of North America* **18**, 141–55 (2004).

14. D. Shoham, Review: Molecular evolution and the feasibility of an avian influenza virus becoming a pandemic strain--a conceptual shift., *Virus Genes* **33**, 127–32 (2006).

15. B. E. Johansson, I. C. Brett, Changing perspective on immunization against influenza., *Vaccine* **25**, 3062–5 (2007).

16. A revision of the system of nomenclature for influenza viruses: A WHO Memorandum, *Bull World Health Organ* , 585–91 (1980).

17. S. Casjens, J. King, Virus assembly, *Annual Review of Biochemistry* , 555–611 (1975).

18. J. C. Richardson, R. K. Akkina, NS2 protein of influenza virus is found in purified virus and phosphorylated in infected cells., *Archives of Virology* **116**, 69–80 (1991).

19. Y. Fujiyoshi, N. P. Kume, K. Sakata, S. B. Sato, Fine structure of influenza A virus observed by electron cryo-microscopy., *The EMBO Journal* **13**, 318–26 (1994).

20. K. G. Murti, P. S. Brown, W. J. Bean, R. G. Webster, Composition of the helical internal components of influenza virus as revealed by immunogold labeling/electron microscopy., *Virology* **186**, 294–9 (1992).

21. J. S. Rossman, G. P. Leser, R. A. Lamb, Filamentous influenza virus enters cells via macropinocytosis., *Journal of Virology* **86**, 10950–60 (2012).

22. W. Chen *et al.*, A novel influenza A virus mitochondrial protein that induces cell death., *Nature Medicine* **7**, 1306–12 (2001).

23. H. M. Wise *et al.*, A complicated message: Identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA., *Journal of Virology* **83**, 8021–31 (2009).

24. M. W. Shaw, P. W. Choppin, R. A. Lamb, A previously unrecognized influenza B virus glycoprotein from a bicistronic mRNA that also encodes the viral

neuraminidase., *Proceedings of the National Academy of Sciences of the United States of America* **80**, 4879–83 (1983).

25. C. M. Horvath, M. A. Williams, R. A. Lamb, Eukaryotic coupled translation of tandem cistrons: Identification of the influenza B virus BM2 polypeptide., *The EMBO Journal* **9**, 2639–47 (1990).

26. J. N. Couceiro, J. C. Paulson, L. G. Baum, Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity., *Virus Research* **29**, 155–65 (1993).

27. T. Ito *et al.*, Molecular basis for the generation in pigs of influenza A viruses with pandemic potential., *Journal of Virology* **72**, 7367–73 (1998).

28. M. Imai *et al.*, Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets., *Nature* **486**, 420–8 (2012).

29. G. N. Rogers, J. C. Paulson, Receptor determinants of human and animal influenza virus isolates: Differences in receptor specificity of the H3 hemagglutinin based on species of origin., *Virology* **127**, 361–73 (1983).

30. M. Matrosovich *et al.*, Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals., *Journal of Virology* **74**, 8502–12 (2000).

31. K. S. Matlin, H. Reggio, A. Helenius, K. Simons, Infectious entry pathway of influenza virus in a canine kidney cell line., *The Journal of Cell Biology* **91**, 601–13 (1981).

32. S. B. Sieczkarski, G. R. Whittaker, Viral entry., *Current Topics in Microbiology and Imunology* **285**, 1–23 (2005).

33. L. H. Pinto, L. J. Holsinger, R. A. Lamb, Influenza virus M2 protein has ion channel activity., *Cell* **69**, 517–28 (1992).

34. G. Neumann, G. G. Brownlee, E. Fodor, Y. Kawaoka, Orthomyxovirus replication, transcription, and polyadenylation., *Current Topics in Microbiology and Immunology* **283**, 121–43 (2004).

35. D. Guilligay *et al.*, The structural basis for cap binding by influenza virus polymerase subunit PB2., *Nature Structural & Molecular Biology* **15**, 500–6 (2008).

36. P. Yuan *et al.*, Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site., *Nature* **458**, 909–13 (2009).

37. G. L. Smith, A. J. Hay, Replication of the influenza virus genome., *Virology* **118**, 96–108 (1982).

38. S. Boulo, H. Akarsu, R. Ruigrok, F. Baudin, Nuclear traffic of influenza virus proteins and ribonucleoprotein complexes, *Virus Research* **124**, 12–21 (2007).

39. B. G. Hale, R. E. Randall, J. Ortı, D. Jackson, The multifunctional NS1 protein of influenza A viruses, *The Journal of General Virology* , 2359–76 (2008).

40. J. S. Rossman, R. A. Lamb, Influenza virus assembly and budding., *Virology* **411**, 229–36 (2011).

41. D. P. Nayak, E. K. W. Hui, S. Barman, Assembly and budding of influenza virus., *Virus Research* **106**, 147–65 (2004).

42. A. P. Schmitt, R. A. Lamb, Influenza virus assembly and budding at the viral budozone., *Advances in Virus Research* **64**, 383–416 (2005).

43. T. Han, W. A. Marasco, Structural basis of influenza virus neutralization., *Annals of the New York Academy of Sciences* **1217**, 178–90 (2011).

44. J. Sui *et al.*, Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses., *Nature Structural & Molecular Biology* **16**, 265–73 (2009).

45. C. Dreyfus *et al.*, Highly conserved protective epitopes on influenza B viruses., *Science* **337**, 1343–8 (2012).

46. D. C. Ekiert *et al.*, Cross-neutralization of influenza A viruses mediated by a single antibody loop., *Nature* **489**, 526–32 (2012).

47. J. R. R. Whittle *et al.*, Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin., *Proceedings of the National Academy of Sciences of the United States of America* **108**, 14216–21 (2011).

48. J. Ma, Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes., *Structure* **13**, 373–80 (2005).

49. M. Lu, B. Poon, J. Ma, A new method for coarse-grained elastic normal-mode analysis., *Journal of Chemical Theory and Computation* **2**, 464–71 (2006).

50. B. K. Poon *et al.*, Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42-Å crystallographic resolution., *Proceedings of the National Academy of Sciences of the United States of America* **104**, 7869–74 (2007).

51. J. Painter, E. A. Merritt, Optimal description of a protein structure in terms of multiple groups undergoing TLS motion., *Acta Crystallographica. Section D, Biological Crystallography* **62**, 439–50 (2006).

52. J. Painter, E. A. Merritt, A molecular viewer for the analysis of TLS rigid-body motion in macromolecules., *Acta Crystallographica. Section D, Biological Crystallography* **61**, 465–71 (2005).

53. F. Zucker, P. C. Champ, E. A. Merritt, Validation of crystallographic models containing TLS or other descriptions of anisotropy., *Acta Crystallographica. Section D, Biological Crystallography* **66**, 889–900 (2010).

54. E. A. Merritt, Expanding the model: Anisotropic displacement parameters in protein structure refinement., *Acta Crystallographica. Section D, Biological Crystallography* **55**, 1109–17 (1999).

55. S.-Y. Lee, A. Lee, J. Chen, R. MacKinnon, Structure of the KvAP voltage-dependent K+ channel and its dependence on the lipid membrane., *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15441–6 (2005).

56. V. V Lunin *et al.*, Crystal structure of the CorA Mg2+ transporter., *Nature* **440**, 833–7 (2006).

57. S. Fath, J. D. Mancias, X. Bi, J. Goldberg, Structure and organization of coat proteins in the COPII cage., *Cell* **129**, 1325–36 (2007).

58. K. Das *et al.*, Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 varian, *Journal of Medicinal Chemistry* **47**, 2550–60 (2004).

59. W. Wang *et al.*, The structure of an open form of an E. coli mechanosensitive channel at 3.45 Å resolution., *Science* **321**, 1179–83 (2008).

60. L. Venkatramani *et al.*, An epidemiologically significant epitope of a 1998 human influenza virus neuraminidase forms a highly hydrated interface in the NA-antibody complex., *Journal of Molecular Biology* **356**, 651–63 (2006).

61. L. Xu, S. D. Benson, R. M. Burnett, Nanoporous crystals of chicken embryo lethal orphan (CELO) adenovirus major coat protein, hexon., *Journal of Structural Biology* **157**, 424–31 (2007).

62. J. Timmins, I. Leiros, S. McSweeney, Crystal structure and mutational study of RecOR provide insight into its mode of DNA binding., *The EMBO Journal* **26**, 3260–71 (2007).

63. K. Viswanathan *et al.*, Glycans as receptors for influenza pathogenesis., *Glycoconjugate Journal* **27**, 561–70 (2010).

64. A. Srinivasan *et al.*, Quantitative biochemical rationale for differences in transmissibility of 1918 pandemic influenza A viruses., *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2800–5 (2008).

65. I. A. Wilson, J. J. Skehel, D. C. Wiley, Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution., *Nature* **289**, 366–73 (1981).

66. J. Shen, B. D. Kirk, J. Ma, Q. Wang, Diversifying selective pressure on influenza B virus hemagglutinin., *Journal of Medical Virology* **81**, 114–24 (2009).

67. J. M. Chen *et al.*, Exploration of the emergence of the Victoria lineage of influenza B virus., *Archives of Virology* **152**, 415–22 (2007).

68. M. N. Matrosovich *et al.*, Probing of the receptor-binding sites of the H1 and H3 influenza A and influenza B virus hemagglutinins by synthetic and natural sialosides., *Virology* **196**, 111–21 (1993).

69. Q. Wang, F. Cheng, M. Lu, X. Tian, J. Ma, Crystal structure of unliganded influenza B virus hemagglutinin., *Journal of Virology* **82**, 3011–20 (2008).

70. Q. Wang, X. Tian, X. Chen, J. Ma, Structural basis for receptor specificity of influenza B virus hemagglutinin., *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16874–9 (2007).

71. Q. Wang, in *Influenza: Molecular Virology*, (2010), pp. 29–52.

72. A. S. Gambaryan, J. S. Robertson, M. N. Matrosovich, Effects of egg-adaptation on the receptor-binding properties of human influenza A and B viruses., *Virology* **258**, 232–9 (1999).

73. J. S. Oxford *et al.*, Direct isolation in eggs of influenza A (H1N1) and B viruses with haemagglutinins of different antigenic and amino acid composition., *The Journal of General Virology* **72**, 185–9 (1991).

74. J. S. Oxford *et al.*, A host-cell-selected variant of influenza B virus with a single nucleotide substitution in HA affecting a potential glycosylation site was attenuated in virulence for volunteers., *Archives of Virology* **110**, 37–46 (1990).

75. J. S. Robertson *et al.*, Alterations in the hemagglutinin associated with adaptation of influenza B virus to growth in eggs., *Virology* **143**, 166–74 (1985).

76. J. S. Robertson *et al.*, The hemagglutinin of influenza B virus present in clinical material is a single species identical to that of mammalian cell-grown virus., *Virology* **179**, 35–40 (1990).

77. T. Saito *et al.*, Antigenic alteration of influenza B virus associated with loss of a glycosylation site due to host-cell adaptation., *Journal of Medical Virology* **74**, 336–43 (2004).

78. G. C. Schild, J. S. Oxford, J. C. de Jong, R. G. Webster, Evidence for host-cell selection of influenza virus antigenic variants., *Nature* **303**, 706–9.

79. N. Ikonen, R. Pyhälä, T. Axelin, M. Kleemola, H. Korpela, Reappearance of influenza B/Victoria/2/87-lineage viruses: Epidemic activity, genetic diversity and vaccination efficacy in the Finnish Defence Forces., *Epidemiology and Infection* **133**, 263–71 (2005).

80. N. Nakagawa, R. Kubota, A. Maeda, T. Nakagawa, Y. Okuno, Heterogeneity of influenza B virus strains in one epidemic season differentiated by monoclonal

antibodies and nucleotide sequences., *Journal of Clinical Microbiology* **38**, 3467–9 (2000).

81. M. Kielian, F. a Rey, Virus membrane-fusion proteins: More than one way to make a hairpin., *Nature Reviews Microbiology* **4**, 67–76 (2006).

82. L. J. Earp, S. E. Delos, H. E. Park, J. M. White, The many mechanisms of viral membrane fusion proteins., *Current Topics in Microbiology and Immunology* **285**, 25–66 (2005).

83. J. M. White, S. E. Delos, M. Brecher, K. Schornberg, Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme., *Critical Reviews in Biochemistry and Molecular Biology* **43**, 189–219 (2008).

84. A. Sapir, O. Avinoam, B. Podbilewicz, L. V Chernomordik, Viral and developmental cell fusion mechanisms: Conservation and divergence., *Developmental Cell* **14**, 11–21 (2008).

85. R. Jahn, Some classic papers in the field of membrane fusion--a personal view., *Nature Structural & Molecular Biology* **15**, 655–7 (2008).

86. J. Chen *et al.*, Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation., *Cell* **95**, 409–17 (1998).

87. J. Chen, J. J. Skehel, D. C. Wiley, N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil., *Proceedings of the National Academy of Sciences of the United States of America* **96**, 8967–72 (1999).

88. P. a Bullough, F. M. Hughson, J. J. Skehel, D. C. Wiley, Structure of influenza haemagglutinin at the pH of membrane fusion.*Nature* **371**, 37–43 (1994).

89. C. M. Carr, P. S. Kim, A spring-loaded mechanism for the conformational change of influenza hemagglutinin., *Cell* **73**, 823–32 (1993).

90. J. A. Gruenke *et al.*, New insights into the spring-loaded conformational change of influenza virus hemagglutinin, *Journal of Virology* **76**, 4456–66 (2002).

91. L. Tamm, Hypothesis: Spring-loaded boomerang mechanism of influenza hemagglutinin-mediated membrane fusion, *Biochimica et Biophysica Acta* **1614**, 14–23 (2003).

92. R. B. Belshe, The need for quadrivalent vaccine against seasonal influenza., *Vaccine* **28**, 45–53 (2010).

93. V. Peltola, T. Ziegler, O. Ruuskanen, Influenza A and B virus infections in children., *Clinical Infectious Diseases* **36**, 299–305 (2003).

94. L. K. Hite, W. P. Glezen, G. J. Demmler, F. M. Munoz, Medically attended pediatric influenza during the resurgence of the Victoria lineage of influenza B virus., *International Journal of Infectious Diseases* **11**, 40–7 (2007).

95. Y. P. Lin, V. Gregory, M. Bennett, a Hay, Recent changes among human influenza viruses., *Virus research* **103**, 47–52 (2004).

96. D. Jackson, R. a Elderfield, W. S. Barclay, Molecular studies of influenza B virus in the reverse genetics era., *The Journal of General Virology* **92**, 1–17 (2011).

97. A. a Vagin *et al.*, REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use., *Acta Crystallographica. Section D, Biological Crystallography* **60**, 2184–95 (2004).

98. R. Blasco, B. Moss, Selection of recombinant vaccinia viruses on the basis of plaque formation., *Gene* **158**, 157–62 (1995).

99. J. Stevens *et al.*, Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus., *Science* **303**, 1866–70 (2004).

100. M. M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding., *Analytical Biochemistry* **72**, 248–54 (1976).

101. J. Stevens *et al.*, Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities., *Journal of Molecular Biology* **355**, 1143–55 (2006).

102. C. C. Wang *et al.*, Glycans on influenza hemagglutinin affect receptor binding and immune response., *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18137–42 (2009).

103. G. Neumann *et al.*, Generation of influenza A viruses entirely from cloned cDNAs., *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9345–50 (1999).

104. K. C. Bradley *et al.*, Analysis of influenza virus hemagglutinin receptor binding mutants with limited receptor recognition properties and conditional replication characteristics., *Journal of Virology* **85**, 12387–98 (2011).

105. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode, *Methods in Enzymology* **276**, 307–326 (1997).

106. H. R. Powell, The Rossmann Fourier autoindexing algorithm in MOSFLM, *Acta Crystallographica Section D Biological Crystallography* **55**, 1690–5 (1999).

107. A. G. W. Leslie, The integration of macromolecular diffraction data., *Acta Crystallographica. Section D, Biological Crystallography* **62**, 48–57 (2006).

108. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments., *Acta Crystallographica. Section D, Biological Crystallography* **67**, 235–42 (2011).

109. P. D. Adams *et al.*, PHENIX: A comprehensive Python-based system for macromolecular structure solution., *Acta Crystallographica. Section D, Biological Crystallography* **66**, 213–21 (2010).

110. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics., *Acta Crystallographica. Section D, Biological Crystallography* **60**, 2126–32 (2004).

111. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models., *Acta Crystallographica. Section A, Foundations of Crystallography* **47**, 110–9 (1991).

112. W. L. Delano, The PyMol molecular graphics system, on World Wide Web http://www.pymol.org (2002).

113. A. T. Brünger *et al.*, Crystallography & NMR system: A new software suite for macromolecular structure determination., *Acta Crystallographica. Section D, Biological Crystallography* **54**, 905–21 (1998).

114. S. C. Lovell *et al.*, Structure validation by Calpha geometry: Phi,psi and Cbeta deviation., *Proteins* **50**, 437–50 (2003).

115. E. A. Govorkova, G. Murti, B. Meignier, C. de Taisne, R. G. Webster, African green monkey kidney (Vero) cells provide an alternative host cell system for influenza A and B viruses., *Journal of Virology* **70**, 5519–24 (1996).

116. S. Milliken, A delay in maturation as a cause of small plaque size with the NM strain of influenza A virus., *The Journal of General Virology* **1**, 189–98 (1967).

117. E. A. Govorkova *et al.*, Selection of receptor-binding variants of human influenza A and B viruses in baby hamster kidney cells., *Virology* **262**, 31–8 (1999).

118. T. Ito *et al.*, Receptor specificity of influenza A viruses correlates with the agglutination of erythrocytes from different animal species., *Virology* **227**, 493–9 (1997).

119. R. Medeiros, N. Escriou, N. Naffakh, J. C. Manuguerra, S. van der Werf, Hemagglutinin residues of recent human A(H3N2) influenza viruses that contribute to the inability to agglutinate chicken erythrocytes., *Virology* **289**, 74–85 (2001).

120. M. A. Larkin *et al.*, Clustal W and Clustal X version 2.0., *Bioinformatics* **23**, 2947–8 (2007).

121. F. S. Cohen, G. B. Melikyan, The energetics of membrane fusion from binding, through hemifusion, pore formation, and pore enlargement, *Journal of Membrane Biology* **199**, 1–14 (2004).

122. A. Jayaraman *et al.*, A single base-pair change in 2009 H1N1 hemagglutinin increases human receptor affinity and leads to efficient airborne viral transmission in ferrets., *PloS One* **6**, e17616 (2011).

123. K. Viswanathan *et al.*, Determinants of glycan receptor specificity of H2N2 influenza A virus hemagglutinin., *PloS One* **5**, e13768 (2010).

124. C. Pappas *et al.*, Receptor specificity and transmission of H2N2 subtype viruses isolated from the pandemic of 1957., *PloS One* **5**, e11158 (2010).

125. V. Y. Lugovtsev, D. F. Smith, J. P. Weir, Changes of the receptor-binding properties of influenza B virus B/Victoria/504/2000 during adaptation in chicken eggs., *Virology* **394**, 218–26 (2009).

126. T. Ito *et al.*, Differences in sialic acid-galactose linkages in the chicken egg amnion and allantois influence human influenza virus receptor specificity and variant selection., *Journal of Virology* **71**, 3357–62 (1997).

127. J. Martín *et al.*, Studies of the binding properties of influenza hemagglutinin receptor-site mutants., *Virology* **241**, 101–11 (1998).

128. D. C. Wiley, J. J. Skehel, The structure and function of the hemagglutinin membrane glycoprotein of influenza virus., *Annual Review of Biochemistry* **56**, 365–94 (1987).

129. S. E. Hensley *et al.*, Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift., *Science* **326**, 734–6 (2009).

130. T. R. Maines *et al.*, Transmission and pathogenesis of swine-origin 2009 A(H1N1) influenza viruses in ferrets and mice., *Science* **325**, 484–7 (2009).

131. J. J. Skehel, D. C. Wiley, Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin., *Annual Review of Biochemistry* **69**, 531–69 (2000).

132. R. Hai, A. García-Sastre, D. E. Swayne, P. Palese, A reassortment-incompetent live attenuated influenza virus vaccine for protection against pandemic virus strains., *Journal of Virology* **85**, 6832–43 (2011).

133. L. J. Mitnaul *et al.*, Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus., *Journal of Virology* **74**, 6015–20 (2000).

134. R. Wagner, M. Matrosovich, H. D. Klenk, Functional balance between haemagglutinin and neuraminidase in influenza virus infections., *Reviews in Medical Virology* **12**, 159–66.

135. R. Xu *et al.*, Functional balance of the hemagglutinin and neuraminidase activities accompanies the emergence of the 2009 H1N1 influenza pandemic., *Journal of Virology* **86**, 9221–32 (2012).

136. R. Lamb, T. S. Jardetzky, Structural basis of viral invasion: Lessons from paramyxovirus F., *Current Opinion in Structural Biology* **17**, 427–36 (2007).

137. S. Martens, H. T. McMahon, Mechanisms of membrane fusion: Disparate players and common principles., *Nature Reviews Molecular Cell Biology* **9**, 543–56 (2008).

138. J. J. Skehel, D. C. Wiley, Coiled coils in both intracellular vesicle and viral membrane fusion., *Cell* **95**, 871–4 (1998).

139. H. S. Choi, J. Huh, W. H. Jo, Electrostatic energy calculation on the pH-induced conformational change of influenza virus hemagglutinin., *Biophysical Journal* **91**, 55–60 (2006).

140. P. S. Rachakonda *et al.*, The relevance of salt bridges for the stability of the influenza virus hemagglutinin., *FASEB Journal* **21**, 995–1002 (2007).

141. E. Borrego-Diaz, M. . Peeples, R. . Markosyan, G. . Melikyan, F. . Cohen, Completion of trimeric hairpin formation of influenza virus hemagglutinin promotes fusion pore opening and enlargement, *Virology* **316**, 234–44 (2003).

142. R. Xu, I. A. Wilson, Structural characterization of an early fusion intermediate of influenza virus hemagglutinin., *Journal of Virology* **85**, 5172–82 (2011).

143. A. Lupas, M. Van Dyke, J. Stock, Predicting coiled coils from protein sequences., *Science* **252**, 1162–4 (1991).

144. D. S. Mueller *et al.*, Histidine protonation and the activation of viral fusion proteins., *Biochemical Society Transactions* **36**, 43–5 (2008).

145. Z. L. Qin, Y. Zheng, M. Kielian, Role of conserved histidine residues in the low-pH dependence of the Semliki Forest virus fusion protein., *Journal of Virology* **83**, 4670–7 (2009).

146. K. Suzuki, T. Yamada, T. Tanaka, Role of the buried glutamate in the alpha-helical coiled coil domain of the macrophage scavenger receptor., *Biochemistry* **38**, 1751–6 (1999).